

Actes des septièmes Journées scientifiques
du réseau de chercheurs
Lexicologie, Terminologie, Traduction

MOTS, TERMES ET CONTEXTES

Sous la direction de :

**Daniel Blampain
Philippe Thoiron
Marc Van Campenhoudt**



Éditions des Archives Contemporaines

as

actualité scientifique



Agence universitaire de la Francophonie

MOTS, TERMES ET CONTEXTES

**Actes des septièmes Journées scientifiques
du réseau de chercheurs
*Lexicologie Terminologie Traduction***

Copyright © 2006 Contemporary Publishing International (CPI). Publié sous licence par Éditions des archives contemporaines et en partenariat avec l'Agence universitaire de la Francophonie (AUF).

Tous droits de traduction, de reproduction et d'adaptation réservés pour tous pays. Toute reproduction ou représentation intégrale ou partielle, par quelque procédé que ce soit (électronique, mécanique, photocopie, enregistrement, quelque système de stockage et de récupération d'information) des pages publiées dans le présent ouvrage faite sans autorisation écrite de l'éditeur, est interdite.

Éditions des archives contemporaines
41, rue Barrault
75013 Paris (France)
Tél.-Fax : +33 (0)1 45 81 56 33
Courriel : info@eacgb.com
Catalogue : www.eacgb.com

ISBN : 2-914610-31-9

Les textes publiés dans ce volume n'engagent que la responsabilité de leurs auteurs. Pour faciliter la lecture, la mise en pages a été harmonisée, mais la spécificité de chacun, dans le système des titres, le choix de transcriptions et des abréviations, l'emploi de majuscules, la présentation des références bibliographies, etc. a été le plus souvent conservée.

Nous tenons à remercier Serge Paulus pour sa contribution aux travaux de prépublication.

Daniel Blampain, Philippe Thoiron et Marc Van Campenhoudt



MOTS, TERMES ET CONTEXTES

**Actes des septièmes Journées scientifiques
du réseau de chercheurs
*Lexicologie Terminologie Traduction***

Bruxelles, Belgique – 8, 9 et 10 septembre 2005

Sous la direction de :

Daniel BLAMPAIN

Institut supérieur de traducteurs et interprètes, Bruxelles, Belgique

Philippe THOIRON

Université Lumière-Lyon II, Lyon, France

Marc VAN CAMPENHOUDT

Institut supérieur de traducteurs et interprètes, Bruxelles, Belgique

COMITÉ SCIENTIFIQUE

Christian BALLIU,
ISTI, Bruxelles (Belgique)
Daniel BLAMPAIN,
ISTI, Bruxelles (Belgique)
Xavier BLANCO,
Université autonome de Barcelone (Espagne)
Anne CONDAMINES,
Université de Toulouse-Le Mirail (France)
Marie-Hélène CORREARD,
présidente d'EURALEX
Thierry FONTENELLE,
Microsoft Natural Language Group, Redmond, WA (États-Unis)
Marie-Claude L'HOMME,
Université de Montréal (Canada)
Salah MEJRI,
Université de Tunis (Tunisie)
Franck NEVEU,
Université de Caen (France)
Philippe THOIRON,
Université Lumière-Lyon II, Lyon (France)

COMITÉ D'ORGANISATION

Daniel BLAMPAIN,
ISTI, Bruxelles (Belgique)
Barbara DEFOUR,
ISTI, Bruxelles (Belgique)
Isabelle GOFFIN,
ISTI, Bruxelles (Belgique)
Michèle LENOBLE-PINSON,
Facultés universitaires Saint-Louis, Bruxelles (Belgique)
Alan SMEKENS,
ISTI, Bruxelles (Belgique)
Marc VAN CAMPENHOUDT,
ISTI, Bruxelles (Belgique)
Philippe THOIRON,
Université Lumière-Lyon II, Lyon (France)

MEMBRES DU COMITÉ DE RÉSEAU LTT

Daniel BLAMPAIN,
ISTI, Bruxelles (Belgique)

Alzouma OUMAROU ISSOUFI,
Université Abdou Moumouni, Niamey (Niger)

Benoît LEBLANC,
Université du Québec à Trois-Rivières, (Canada-Québec)

Salah MEJRI,
Université de la Manouba, Tunis (Tunisie)

Maria Teresa RIJO DA FONSECA LINO,
Universidade Nova de Lisboa, Lisbonne (Portugal)

Philippe THOIRON,
Université Lumière-Lyon II, Lyon, (France) coordonnateur du Réseau

André CLAS,
Université de Montréal, Montréal (Canada-Québec), président d'honneur

APPEL À COMMUNICATIONS



JOURNÉES SCIENTIFIQUES

Réseau de chercheurs *Lexicologie, terminologie et traduction*
Langue française, Francophonie et Diversité linguistique
AGENCE UNIVERSITAIRE DE LA FRANCOPHONIE

Appel à communications
7^{es} Journées scientifiques
« **Mots, termes et contextes** »
Bruxelles, 8 au 10 septembre 2005

DOCUMENT DE PRÉSENTATION (extraits)

1. Contexte

Le réseau de chercheurs *Lexicologie, Terminologie et Traduction* organise ses septièmes Journées scientifiques qui se tiendront à Bruxelles (Belgique) en septembre 2005. Ces Journées, organisées en collaboration avec l'Institut supérieur de traducteurs et interprètes (ISTI), auront pour titre « Mots, termes et contexte », et seront précédées, les 6 et 7 septembre, des Journées de formation « Informatique et contextes ».

Plus d'informations sur les Journées de formation et comment y participer, à partir de la page Internet du programme *Langue française, Francophonie et Diversité linguistique* :

www.auf.org/programmes/programme1/

Le réseau *Lexicologie, Terminologie et Traduction* de l'AUF

Le réseau *Lexicologie, Terminologie et Traduction*, créé en 1988, a pour objectif :

- de soutenir, encourager, promouvoir les travaux de recherche, la publication et la formation dans les domaines de compétence du réseau en assurant une modernité justifiée et un appui technologique pour faire face aux développements humains, sociaux, politiques et économiques de la réalité francophone des divers pays de la francophonie ;
- d'assurer la production d'outils de référence en langue générale et en langue de spécialité : lexiques, dictionnaires, bases de données, outils d'interprétation de données, traduction, traductologie, traductique.

Les Journées scientifiques de Bruxelles seront les septièmes, les autres s'étant déroulées à Fès (1989), à Mons (1991), à Montréal (1993), à Lyon (1995), à Tunis (1997) et à Beyrouth (1999).

Pour plus d'informations sur le réseau de chercheurs *Lexicologie, Terminologie et Traduction* :

www.ltt.auf.org

L'Institut supérieur de traducteurs et interprètes (ISTI)

L'ISTI fait partie de la **Haute École de Bruxelles** depuis 1996 (présidence de la Haute école : Mme M. COESSENS; direction de la section traduction et interprétation : M. F. DE LAET). L'ISTI est une institution universitaire francophone membre de l'Agence universitaire de la Francophonie.

Le centre de recherche **TERMISTI** (direction : D. BLAMPAIN) est rattaché à l'ISTI. Ses activités sont essentiellement orientées vers l'étude des vocabulaires spécialisés et de leur description à l'aide d'outils logiciels. Bénéficiant des compétences scientifiques et professionnelles du corps professoral de l'institut, il sert de centre de référence aux étudiants qui préparent un mémoire de fin d'études (bac + 4 ou bac + 5) consacré aux langues spécialisées.

Les activités de TERMISTI se développent dans le cadre des besoins de l'ISTI, de la Haute École de Bruxelles et de programmes de recherche francophones ou internationaux.

Pour plus d'informations :

www.heb.be/isti/
www.termisti.refer.org
daniel.blampain@termisti.be

2. Objectifs des Journées scientifiques

Les **septièmes Journées scientifiques** du réseau *Lexicologie, Terminologie, Traduction*, prévues du 8 au 10 septembre 2005, à Bruxelles sur le thème « **Mots, termes et contextes** », réuniront les experts, universitaires et chercheurs travaillant dans plusieurs pays de la Francophonie.

Ces Journées permettront de faire le point sur les nouvelles dynamiques qui régissent les rapports entre les langues. Une grande attention sera également portée sur les communications développant des approches comparatives dans le temps ou entre pays.

Les Journées scientifiques seront précédées de **deux Journées de formation, les 6 et 7 septembre 2005**, sur le thème « **Informatique et contextes** », avec les linguistes et informaticiens du Laboratoire de linguistique informatique de l'Université Paris XIII.

3. Thématiques et questions à débattre

Une des problématiques devenues dominantes aujourd'hui en lexicologie, en terminologie et en traduction est celle des **contextes**. La linguistique de *corpus*, née de la capacité actuelle de traiter de grandes quantités de documents, remet en cause les frontières entre disciplines. Tout se joue dans l'analyse de *corpus* textuels et non plus simplement dans la pragmatique d'une phraséologie. Le contexte cesse d'avoir valeur d'illustration et représente l'élément fondamental de la construction du sens, notamment dans le traitement automatique des langues.

Le traducteur demande des *corpus* de référence fiables en matière de langues de spécialité. Il est à la recherche des liens pertinents entre le texte de *corpus* et le texte à traduire. La circulation des textes au-delà des domaines doit permettre de recentrer son activité sur les contraintes communicationnelles.

Le terminologue est amené à remettre en cause le système conceptuel propre à un domaine, le pouvoir de structuration du domaine et le logicisme ambiant au profit de pratiques textuelles réelles. Il se trouve confronté par les macrocontextes à la variabilité des terminologies, à une autre typologie des textes et à une autre acception de la normalisation.

Le lexicologue et le lexicographe doivent analyser les implications de cette linguistique textuelle, redéfinir la notion de proximité dans l'utilisation de leurs outils informatiques et poser, selon de nouveaux principes, le mot, l'exemple, la citation.

La priorité accordée au contexte donne naissance à de nouvelles procédures et à de nouvelles considérations théoriques, au carrefour des différentes disciplines. Lesquelles ? Aux lexicologues, aux traducteurs, aux terminologues et aux enseignants de tenter d'y répondre au cours des Journées scientifiques de Bruxelles.

LISTES DES AUTEURS

Alexandru Maria-Cristina, laboratoire dynamiques sociolangagières, Université de Rouen, France

Anastassiadis-Syméonidis Anna, département de philologie, section de linguistique, Université Aristote de Thessaloniki, Thessaloniki, Grèce

Arnaud Pierre J-L., CRTT, Université Lumière, Lyon, France

Ayewa Noël Kouassi, département des sciences du langage, Université de Cocody, Abidjan, Côte d'Ivoire

Balliu Christian, ISTI, Bruxelles, Belgique

Baptista Jorge, facultade de ciências humanas e sociais, DLCM, Universidade do Algarve, Portugal

Barna Corina Georgeta, faculté de médecine vétérinaire, Université de sciences agricoles et médecine vétérinaire du Banat de Timisoara, Roumanie

Bellynck Valérie, Institut national polytechnique de Grenoble, Grenoble, France

Ben Hariz Sondos, Université de la Manouba, Tunis, Tunisie

Bertaccini Franco, SSLMIT Forli, Université de Bologne, Italie

Bertels Ann, Institut des langues vivantes, K. U. Leuven, Belgique

Blaise Jean-Yves, UMR – CNRS/MCC 694 MAP, Marseille, France

Boitet Christian, GETA-CLIPS-IMAG, Université Joseph Fourier, Grenoble, France

Catala Dolors, département de philologie française et romane, Université autonome de Barcelone, Espagne

Contente Madalena, centro de linguística-UNL, DEM-Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Portugal

Costa Rute, centro de linguística da Universidade Nova de Lisboa, Portugal

Daille Béatrice, LINA FRE CNRS 2729/LLING EA 3827, Université de Nantes, France

Desmet Isabel, département de portugais et Certal, UFR V, Université Paris VIII, France

Diakité Mamadou, Centre de linguistique appliquée de Dakar, CLAD, Sénégal

Dubreil Estelle, Lina FRE CNRS 2729 / LLING EA 3827, Université de Nantes, France

Dudek Iwona, UMR – CNRS/MCC 694 MAP, Marseille, France

Durieux Christine, Maison de la recherche en sciences humaines, Université de Caen, France

Dussart André, ISTI, Bruxelles, Belgique

Elimam Abdeljlil Abdou, ENSET, Oran, Algérie

Escartin Arilla Ana, departamento de filología francesa, Universidad de Zaragoza, Teruel, Espagne

Fafiotte Georges, GETA, laboratoire CLIPS-IMAG, UJF, CNRS, Institut national polytechnique de Grenoble, France

Fontanet Mathilde, ETI, Université de Genève, Suisse

Gaudin François, laboratoire dynamiques sociolangagières, Université de Rouen, France

Goffin Roger, Université libre de Bruxelles, Belgique

González Rodríguez Antonio, département de traduction et interprétation, Universidad Europea de Madrid, Espagne

Ilunga Ntumba Willy, département des lettres françaises, Université de Kinshasa, République démocratique du Congo

Jacquemin Bernard, Institut des sciences cognitives, CNRS UMR 5015, Bron, France

Jacques Marie-Paule, ERSS, Université Toulouse-le Mirail, France

Kenwright John, Institut national polytechnique de Grenoble, Grenoble, France

Kerremans Koen, département de linguistique appliquée, Erasmushogeschool Brussel, Belgique

Lerat Pierre, LLI, Université Paris XIII, France

Leroyer Patrick, École des hautes études commerciales de Århus, Århus, Danemark

Lino Teresa, Centre de linguistique, Université nouvelle de Lisbonne, Portugal

Massoumou Omer, département de langue et littérature françaises, Université Marien Ngouabi, Brazzaville, République du Congo

Mathieu-Colas Michel, laboratoire de linguistique informatique, Université de Paris XIII, France

Matteuci Alessandra, SSLMIT Forli, Université de Bologne, Italie

Mejri Salah, LLI, Université Paris XIII, France

Mejri Soumaya, IAE Basse-Normandie, Université de Caen, France

Merten Pascaline, ISTI, Bruxelles, Belgique

Messaoudi Leila, département de langue et littérature française, UFR de sociolinguistique appliquée, Université de Kénitra, Maroc

Mokhtari Abdelkrim, faculté des lettres, Université Ibn Tofail, Kénitra, Maroc

Muraille, Paul, ISTI, Bruxelles, Belgique

Mustafa el Hadi Widad, UFR IDIST/CERSATES, UMR 8529, Université Lille III, France

Ndiaye Modou, faculté des lettres, Université Cheikh Anta Diop, Dakar, Sénégal

Neveu Franck, département de sciences du langage, Université de Caen, France

Nzesse Ladislas, faculté des lettres et sciences humaines, Université de Dschang, Cameroun

Ouerhani Béchir, unité de recherche 00/UR/0201, Université de la Manouba, Béja, Tunisie

Phan Huy Khánh, Université de Danang, Danang, Viêt Nam

Ravaoarimalala Elisabeth, Alliance française d'Antananarivo, Madagascar

Sader Feghali Lina, Centre d'études et de recherches en terminologie arabe, École de traducteurs et interprètes de Beyrouth, Liban

Said Mosbah, ISET'COM, RLM, Tunis, Tunisie

Sanogo Mamadou Lamine, Centre national de la recherche scientifique et technologique, Ouagadougou, Burkina Faso

Selosse Philippe, GRAC / UMR 5037, Université Lumière-Lyon II, France

Silva Raquel, Centre de linguistique, Universidade Nova de Lisboa, Portugal

- Simonsen Henrik K.**, École des hautes études commerciales de Copenhague, Frederiksberg, Danemark
- Sliwa Dorota**, Institut de philologie romane, Katolicki Uniwersytet Lubelski, Lublin, Pologne
- Surmont (de) Jean-Nicolas**, Bruxelles, Belgique
- Temmerman Rita**, département de linguistique appliquée, Erasmushogeschool Brussel, Belgique
- Thierno Cissé**, UCAD, Dakar, Sénégal
- Torterat Frédéric**, Université Paris-Sorbonne/Université d'État d'Haïti, Paris, France
- Tosheva Aneta**, section de langues étrangères, Université de médecine de Plovdiv, Bulgarie
- Tsahnaki Olympia**, département de philologie, section de linguistique, Université Aristote de Thessaloniki, Thessaloniki, Grèce
- Trurnit Verbic Renate**, Nice-Sophia Antipolis, LILLA-EA 2140 /CRDL, Valbonne, France
- Van Campenhoudt Marc**, ISTI, Bruxelles, Belgique
- Van der Veken Anneleen**, Linguistique africaine, Université libre de Bruxelles/Musée royal de l'Afrique centrale, Tervuren, Belgique
- Vandervoort Veerle**, département de linguistique appliquée, Erasmushogeschool Brussel, Belgique
- Vangehuchten Lieve**, département de communication internationale de gestion, section d'espagnol, Université d'Anvers, Belgique
- Vaxelaire Jean-Louis**, CIEL, Université Paris VII, France
- Vergely Pascale**, ERSS, Université de Toulouse-Le Mirail, France
- Vo Trung Hung**, Université de Danang, Danang, Viêt Nam
- Zimina-Poirot Maria**, CLA2T Syled, Université de la Sorbonne nouvelle-Paris III, France

SÉANCE D'OUVERTURE

Allocution d'ouverture



Michèle Gendreau-Massaloux
recteur de l'Agence universitaire de la Francophonie

*Cher Daniel Blampain,
Monsieur le commissaire général aux relations internationales,
Madame la directrice générale du Service de la langue française,
Madame la directrice-présidente de la Haute École de Bruxelles,
Monsieur le directeur de l'Institut supérieur de traducteurs et interprètes,
Monsieur Philippe Thoiron, coordonnateur du réseau Lexicologie, Terminologie,
Traduction,
Mes chers collègues,*

Je suis heureuse de retrouver parmi vous, parmi les composantes les plus vivantes de la Francophonie scientifique, celle qui fait de la langue française un outil pour la recherche, et de la recherche un levier pour le développement. Demain plus encore qu'aujourd'hui, il faudra reconnaître la capacité de la recherche, non seulement à créer des emplois, mais à représenter, culturellement, une des raisons pour les jeunes de travailler dans un monde plus généreux, plus ouvert et qui partage le savoir.

Je suis heureuse, aussi, de me trouver accueillie ici par des institutions qui portent très haut le flambeau de la recherche universitaire et qui, en même temps, sont intégrées à des ensembles politiques qui la soutiennent. La Communauté française de Belgique, riche de réseaux comme l'Agence universitaire de la Francophonie, a mis l'accent sur les « réseaux éducatifs ». Les mots ont leur importance. Il me semble que c'est à l'honneur de l'ISTI d'avoir fait l'effort, avec l'Agence universitaire, d'accueillir et même de financer en partie ces manifestations comme un appui à leurs réseaux : elles prouvent que l'exigence de rigueur scientifique s'accompagne d'un grand appel au partage du savoir et à la formation des jeunes de tous les pays. Dans les « journées » organisées par des réseaux de chercheurs comme le vôtre, et soutenues par des institutions comme l'Isti, je remarque le souci

de s'ancrer sur la richesse de propositions locales, en donnant lieu à ce que l'on appelle souvent, de manière remarquable, les « journées d'animation régionale ».

Je remarque aussi qu'au lieu de considérer comme beaucoup d'universitaires votre réseau comme un ensemble clos où l'on se contenterait de se retrouver entre soi à intervalles réguliers, l'événement est mis à profit pour la formation des jeunes chercheurs. J'apprécie particulièrement que les séminaires soient conduits autour des technologies de l'information, qui représentent pour toutes les disciplines un apport non négligeable, et en particulier pour les vôtres. Elles permettent aussi de faire œuvre de partage en élevant le niveau de ceux qui accompagnent votre démarche aujourd'hui à Bruxelles.

Si j'avais à traiter le sujet que vous vous êtes proposé, « Mots, termes, contextes », je m'interrogerais sur une notion que connaît bien, en termes politiques, la Communauté française de Belgique : je choiserais comme mots, pour les situer dans leur contexte, « identité » et « territoire ». On peut en effet les analyser à la fois selon le contexte politique, celui d'une communauté dans un pays fédéral, ou encore selon le contexte universitaire dans lequel se situent les institutions où vous-mêmes enseignez ou étudiez, et en troisième lieu selon le contexte d'une organisation de coopération multilatérale pour le développement comme l'Agence universitaire de la Francophonie. Ceux qui se trouvent dans une université classique, ou dans une autre grande école de qualité scientifique reconnue, sont confrontés à l'indispensable nécessité d'affirmer ce qu'ils sont et de montrer l'excellence de ce qu'ils font, avec les forces qui sont les leurs.

En contexte multilatéral, en revanche, la notion d'identité se conjugue nécessairement avec la notion d'ouverture. L'une ne va pas sans l'autre. La dernière fois que j'ai eu le plaisir de me rendre à des journées organisées par le réseau LTT, selon ce vocabulaire que je n'aime guère parce que les sigles me semblent assez largement contradictoires avec les langues, j'étais accompagnée par un philosophe, qui n'est plus de ce monde. Il s'appelait Jacques Derrida et le colloque qui lui était consacré s'intitulait « Le Passage des frontières ». Lorsque l'on réfléchit à la notion d'identité et de territoire en termes de développement multilatéral, si l'on veut mettre en pratique la solidarité des États et suivre le mouvement qui amène la Francophonie des États, comme elle l'a fait à Ouagadougou, à réaffirmer certains axes autour desquels elle est lisible et se propose de demander aux autres opérateurs une programmation en phase avec ses orientations, on se rend compte que ni la notion d'identité ni la notion de territoire ne peuvent se définir simplement comme des exigences d'identification ou de territorialisation.

Il n'y a pas d'identité qui se construise sans se confronter à une altérité. Il n'y a pas de territoire qui se délimite sans que des ouvertures doivent y être aménagées afin de permettre l'apport d'autres territoires. Il ne peut y avoir de réseaux sans un contact permanent, dans le respect de leur spécificité puisqu'il faut évidemment que chacun garde et protège son identité, avec ce que peuvent leur apporter d'autres réseaux. La seule règle qui garantisse non seulement la tolérance et le dialogue, mais aussi la science, c'est le croisement des identités et le passage des frontières entre les territoires.

Je vous sais engagés sur cette piste, comme la Communauté française de Belgique, avec le double souci de conserver votre identité et d'accueillir des propositions en provenance d'autres territoires, d'autres composantes de la francophonie scientifique.

Toute institution vivante réécrit en permanence les lignes de son avenir. Vous le faites par l'action des coordonnateurs de vos réseaux, par celle des chercheurs qui sont auprès de vous, non seulement dans le comité, en train de penser la science, mais aussi dans des lieux plus isolés, en liaison constante avec ce réseau qui les soutient et les encourage. Penser la science c'est l'écrire, comme vous en avez l'occasion par les communications de ces journées scientifiques ; c'est aussi essayer d'attirer et de rallier ceux qui, ne connaissant pas encore la réalité de votre travail, souhaitent dialoguer avec d'autres composantes des institutions universitaires et de recherche qui vous entourent, et dont l'Agence universitaire de la Francophonie est fière. Je voudrais en particulier, et en terminant, remercier encore celle qui nous accueille et qui me semble de ce point de vue, avec d'autres, exemplaire.

Je vous remercie.

Allocution d'ouverture



Philippe Thoiron
coordonnateur du réseau LTT

Mesdames, Messieurs,

Je souhaite tout d'abord remercier l'ISTI et la HEB, qui ont accepté d'accueillir ces Journées LTT 2005. Nous avons pu ainsi organiser les Journées de formation et les Journées scientifiques sur un même lieu, ce qui facilite naturellement les rencontres et les échanges entre participants. Je veux aussi témoigner la gratitude du comité de réseau LTT à Madame la rectrice de l'AUF, Madame la directrice générale du Service de la langue française, Madame la directrice-présidente de la HEB, Monsieur le commissaire général aux relations internationales, et Monsieur le Directeur de l'ISTI qui ont bien voulu participer à cette séance d'ouverture. Il va de soi que rien de tout cela n'aurait été possible sans le dévouement et la maîtrise dont ont témoigné au cours de ces derniers mois notre collègue Daniel Blampain, membre du comité LTT, et toute son équipe.

C'est avec un plaisir tout particulier que le coordonnateur que je suis renoue avec la tradition du couplage des Journées de formation et Journées scientifiques inauguré à Montréal en 1993 et renforcé à Lyon en 1995. Si ce couplage est à nouveau possible c'est aussi parce que, après une accalmie que certains ont trouvée trop longue, nous avons pu convoquer ces 7^{es} Journées scientifiques. Chacun ici sait que cette accalmie a été concomitante et consécutive au passage de l'AUF à l'AUF. Un nouveau comité et un nouveau coordonnateur ont eu la charge de mettre un terme à cette mise en sommeil des actions du réseau LTT.

La nature même du réseau LTT, le plus ancien des réseaux Langues et celui dont les domaines de compétences sont centraux pour une Francophonie soucieuse du dialogue des langues et des cultures, conduisait inévitablement à des préoccupations relatives au maintien de sa spécificité, elle-même liée à son intégrité. Dans le même temps, sont apparues les nécessités de collaborations plus étroites avec d'autres composantes, au sein de l'AUF aussi bien qu'à l'extérieur. Nous nous sommes acquittés des missions qui nous ont été confiées en participant aux Journées communes de Beyrouth en 2001 et de Ouagadougou en 2004. Nous avons aussi remis sur le métier un travail plus spécifique à notre réseau. Dans le cadre de la journée d'animation régionale de Hammamet en 2004 et, bien sûr, dans le cadre de ces 7^{es} Journées scientifiques. Il

me semble que nous remplissons ainsi, de manière cohérente, un rôle auquel notre comité tient beaucoup. À savoir, favoriser l'ouverture du réseau LTT, dans ses domaines de compétence, ses objectifs et ses modes d'action, en préservant sa visibilité et sa spécificité.

Ces Journées scientifiques se placent donc dans une longue tradition fondée à Fès en 1989, poursuivie ensuite à Mons (la Belgique déjà), Montréal, Lyon, Tunis et Beyrouth. Les Journées scientifiques LTT ont rassemblé, et rassemblent encore cette fois, des chercheurs qui viennent de nombreux pays et travaillent, que ce soit au sein de LTT ou d'autres réseaux Langue, sur des thèmes relatifs au dialogue des cultures à travers des projets essentiellement bilingues ou plurilingues. Les actes de ces Journées scientifiques ont tous été publiés et la qualité de ces travaux, qui a été reconnue par les communautés compétentes, a permis de valoriser la recherche de nos collègues du Sud, y compris dans les domaines neufs de l'ingénierie linguistique.

Le choix de la thématique de ces Journées scientifiques et celle des Journées de formation montre d'ailleurs clairement l'importance que nous accordons à ces domaines neufs. Vous y voyez associés les mots, les termes, les textes, les contextes et bien sûr l'informatique.

Les évolutions théoriques et applicatives liées aux nouvelles technologies tendent à accroître les écarts Nord-Sud et augmentent parfois les réticences des laboratoires du Nord à l'égard d'un partenariat avec le Sud. Le différentiel technoscientifique, qu'il soit supposé ou réel, doit être pris en compte si l'on souhaite que la mise en place d'actions Nord-Sud, clé de voûte de notre ensemble, ne se limite pas à un vœu pieux. C'est pourquoi nous pensons que ce que j'appellerais volontiers « le modèle Hammamet » est viable et efficace. L'économie générale des journées d'animation régionale a été à cette occasion repensée grâce à un interfaçage fructueux entre recherche et formation par l'intermédiaire des Actions de recherche partagée (les ARP). C'est aussi pourquoi nous avons choisi pour les Journées de formation une thématique qui soit naturellement parente de celle des Journées scientifiques mais qui soit également directement ancrée sur les préoccupations et les besoins de nos collègues du Sud et de leurs étudiants.

Si j'en juge par le nombre de demandes qui sont parvenues au comité il me semble que nos choix ont été compris et acceptés. S'agissant plus spécialement des Journées scientifiques qui vont commencer dans quelques instants, je tiens à dire d'emblée que le nombre élevé de propositions de communications (plus de 90, issues de plus de 30 pays différents) a permis au comité scientifique de ces 7^{es} Journées scientifiques de conserver un grand nombre de communications de qualité. Nous nous en réjouissons évidemment, même si cela a contraint le comité d'organisation à prévoir des sessions parallèles. Nous nous efforcerons, avec le concours des présidents de séance que je remercie d'avance pour leur aide, de garder entre les horaires des communications le parallélisme annoncé.

Dans l'ambiance chaleureuse et conviviale qui caractérise ce lieu et témoigne des efforts de nos hôtes, il ne me reste plus qu'à vous souhaiter de passer trois journées

agréables et fructueuses. Au nom du comité LTT je compte sur vous pour nouer et resserrer les liens qui contribuent à renforcer les valeurs scientifiques de la francophonie et de notre agence universitaire en permettant à notre réseau de maintenir et d'accroître sa vitalité et son dynamisme. Bon travail à tous !

Allocution d'ouverture



Daniel Blampain
directeur de TERMISTI

Mesdames, Messieurs,

Le centre de recherche en terminologie de l'Institut supérieur de traducteurs et interprètes, TERMISTI, dont le logo marque ces journées, a mis tout en œuvre, avec l'Institut et la Haute École, pour que vos travaux se déroulent dans les meilleures conditions possibles. Ils ont été aidés par le Commissariat général aux relations internationales, le Service de la langue française de la Communauté française de Belgique, le Fonds national de la recherche scientifique et la ville de Bruxelles.

Symboliquement, scientifiquement, pratiquement, tout est réuni pour que tout aille pour le mieux. Mais tout est encore à jouer.

Symboliquement, le logo que vous avez sous les yeux et sous la main, sur et dans votre porte-documents, je l'ai créé il y a plus de vingt ans pour mon centre de recherche. Il évoque l'Atomium de Bruxelles, il évoque le passé et l'avenir. L'Atomium, dû à l'ingénieur André Waterkeyn, évoque le concept de l'atome et de son utilisation pacifique. Construit pour la première exposition universelle de 1958, année de création de l'ISTI, aux lendemains de la guerre, pour marquer l'engagement dans l'utilisation pacifique de l'atome, il représente un cristal élémentaire de fer, à l'échelle de ses atomes, agrandi cent soixante milliards de fois. Neuf grandes sphères reliées par des tubes, disposées selon la configuration du système cubique centré, mettent en évidence, de façon monumentale, ce qui se passe à l'échelle de l'infiniment petit. 1958 signifie aussi, par cette exposition universelle, une campagne tout à fait extraordinaire pour la promotion des langues étrangères en Belgique. « Soyez prêts pour 1958, étudiez les langues » fut le slogan choisi pour les affiches. Vingt-quatre mille interprètes bénévoles furent présents sur le site. Aujourd'hui, cinquante ans plus tard, la préoccupation de la Francophonie est de situer le français parmi le multilinguisme. Bruxelles, l'ISTI, lieux par excellence du multilinguisme.

Dans la recherche terminologique, nous nous préoccupons des langues et de l'analyse de l'infiniment petit, des termes, mais notre pouvoir d'intervention dans la création des concepts, notamment en néologie, est important. De là, la présence des deux rectangles à la droite du logo, qui créent l'inattendu par rapport à l'harmonie des sphères. Le tout pouvant être considéré comme le symbole des réseaux notionnels sur lesquels ont porté souvent nos recherches.

Scientifiquement, lorsque le comité *Lexicologie, Terminologie, Traduction* de l'Agence universitaire de la Francophonie a accepté ma proposition de traiter les contextes au cours de ces journées scientifiques, c'était aller de l'infiniment petit à l'infiniment grand. Aujourd'hui, par notre capacité à gérer des *corpus*, nous sommes arrivés à un tournant dans la recherche. Nous reconnaissons une forme de complexité dans la catégorisation des unités linguistiques, nous localisons les interactions entre les unités et l'environnement, mais comment spécifier ces interactions ? Nous sommes contraints de déterminer des paliers de contextualité (immédiate, proche, lointaine) et de nous tenir à distance du romantisme des totalités (*corpus* non étiquetés), de l'immaculée perception ou des curiosités, devenues attendrissantes, que représentent les exemples forgés. En d'autres mots, comment les contextes permettent-ils aujourd'hui la construction du sens ? Que de remises en cause ! La pragmatique de la phraséologie, les frontières du texte à traduire, les frontières des disciplines, le logicisme des démarches terminologiques, voire linguistiques en général, la conception idéalisée des domaines du savoir. Devant ces textes qui prolifèrent et se bousculent dans tous les sens, nous devons, en toute modestie et en toute clarté, définir de nouveaux concepts, de nouvelles approches descriptives, de nouvelles procédures pour le lexicologue, le terminologue et le traducteur. En cela, le travail en réseau comme celui de LTT, qui met en communication le plus grand nombre de chercheurs, venus de tous les horizons de la francophonie, n'est pas seulement souhaitable, il est indispensable, il est vital pour l'avenir de la langue française.

Pratiquement, vos séances de travail se dérouleront dans trois salles différentes, en parallèle, selon des thématiques générales que vous percevrez en filigrane des titres de communications. Nous travaillons en parallèle tellement vous avez répondu nombreux à notre appel. Le comité scientifique a eu fort à faire, mais se réjouit de l'enthousiasme et de l'intérêt suscité par ces journées.

Enfin, j'ose espérer qu'à la fin de nos travaux, le logo d'aujourd'hui, vide de tout mot ou de toute lettre, pourra comporter le mot réussite dans un de ses rectangles, mieux j'ose espérer que *mots, termes et contextes* ne seront plus évoqués sans la référence à Bruxelles 2005.

COMMUNICATIONS

PHÉNOMÈNES DE SOUS-SPÉCIFICATION SÉMANTIQUE REPRÉSENTATION DES UNITÉS LEXICALES ET PRISE EN COMPTE DU CONTEXTE

Pierre J.-L. Arnaud
CRTT
Université Lumière
Lyon, France

Résumé : Trois phénomènes linguistiques ou psycholinguistiques, certains lapsus, la métonymie et la composition nominale, manifestent une variété de liens sémantiques implicites considérable et difficile à classer entre, respectivement, cible et erreur, source et cible, et N1 et N2. Après avoir exposé les problèmes, on envisage les processus et les représentations des unités lexicales qui pourraient permettre de les expliquer, en rendant compte du contexte là où celui-ci joue un rôle.

Mots-clés : lapsus, métonymie, noms composés, représentations lexicales.

1. INTRODUCTION

Je voudrais rapprocher ici divers phénomènes langagiers qui tous impliquent d'une part la présence implicite de liens sémantiques parfois complexes entre paires de mots, et d'autre part une grande variété dans la nature de ces liens, qui en rend l'inventaire difficile.

2. LAPSUS

Les lapsus par substitution sémantique de mots sont un premier exemple. Parmi les nombreuses catégories de lapsus, certains ont pour résultat la substitution d'un mot-erreur au mot-cible, sans qu'on puisse incriminer une substitution de segments. Les substitutions de mots se répartissent en substitutions à cause a) formelle ; b) contaminationnelle (le nom d'un élément du contexte d'énonciation remplace la cible) ; et c) sémantique, avec des catégories mixtes.

Les 322 substitutions sémantiques de mes données (Arnaud 1999) comprennent 231 lapsus non-antonymiques. Ceux-ci sont presque exclusivement du type co-hyponymique ou co-méronymique au sens large de ces termes, et jamais hyperhyponymique ou synonymique, ce qui confirme ce qu'on trouve dans d'autres langues. L'exemple suivant est un cas indiscutable de co-hyponymie :

[1] [en apportant un plat de cerises] Comme dessert, maintenant, nous avons des fraises//non !

En effet, on peut intégrer cible et erreur à un énoncé test :

Cerises et fraises sont des fruits rouges

Un tel test indique l'existence d'un hyperonyme, et c'est dans ce cas qu'il est légitime de parler de co-hyponymie. Mais d'autres substitutions sont plus problématiques à cet égard :

[2] [en expliquant à un enfant qu'il faut se brosser les dents immédiatement après les repas]

C'est le meilleur moment parce que c'est là que les insectes//les insectes ! les microbes se mettent à attaquer.

L'énoncé-test de la co-hyponymie donnerait ici un résultat peu heureux. Et dans l'exemple suivant, il n'est simplement pas possible de trouver un hyperonyme acceptable :

[3] [en découvrant les douches d'un camping]
On va prendre des bains.

Or, le lien sémantique cible-erreur est subjectivement évident. Ce genre de cas crée une gêne manifeste dans la littérature, où Garrett (1980) parle d'un « nombre substantiel de cas où la relation est plutôt vague ». Hotopf (1980) signale dans ses données 45 % de co-hyponymes et 24 % de mots en relation « approximativement hyponymique ». Dans un article postérieur (Hotopf, 1983), cet auteur emploie le terme générique de « *semantic group slips* », les co-hyponymes allant de « *tightly constricted* » à « *loosely constricted* ». Harley (1993) parle de « coordonnés de même champ sémantique ».

Voici encore trois exemples qui illustrent la difficulté de catégorisation :

[4] – Vous n'avez pas eu froid ?
– Oh, tu sais, avec un anorak/un édredon...

[5] C'est ma montre/ma montre... non, ma bague, qui vient de me lâcher.

[6] Et tes parents, ils ont toujours leur tente/leur tombe toute neuve ?

S'il n'y a pas d'hyperonyme pour respectivement anorak et édredon, montre et bague, et tente et tombe, il n'est pas pour autant souhaitable de se replier sur des champs sémantiques invoqués ad hoc, d'autant plus qu'on sait les reproches qui ont été adressés en leur temps aux champs sémantiques (v. p. ex. Lehrer 1974 : 66 ; Wierzbicka 1985 : 74). La ressemblance sémantique se laisse certes analyser informellement : anoraks et édredons sont des artefacts textiles chauds, rembourrés et au piqué particulier. Montres et bagues sont des artefacts portés autour de parties

du corps proches, annulaires si l'on prend en compte le bracelet, plutôt coûteux, etc. Tentés et tombés sont des abris pour des corps couchés ¹, etc. Mais les problèmes liés à la notion de champ sémantique nous interdisent de voir dans ces éléments de ressemblance des sèmes, unités d'un modèle rigoureusement établi, et de considérer le problème comme réglé.

3. MÉTONYMIES

La métonymie pose le même problème. Elle est multiple et envahissante, et passe facilement inaperçue :

[7] Il parle espagnol. [= parle et comprend]

Elle peut affecter les prédicats, comme dans l'exemple suivant :

[8] Mamie Nova. Les mamies ne lui disent pas merci.

Les métonymies peuvent s'empiler les unes sur les autres :

[9] Oscar du meilleur acteur

et aussi coexister avec une métaphore :

[10] Les cheveux courts, la vingtaine, il portait le bouc rasé de près, à la mode

Les problèmes surviennent quand on cherche à classer une collection de métonymies. Certes, beaucoup d'entre elles, comme les deux suivantes, se laissent regrouper aisément :

[11] deux escadrons de 72 sabres [cavaliers de la Garde républicaine]

[12] une baguette très recherchée [chef d'orchestre]

Il s'agit ici de la catégorie bien connue l'instrument/accessoire de l'activité (source) pour l'individu qui l'utilise (cible).

Il existe comme celle-ci des catégories puissantes qui par analogie facilitent la création de nouvelles occurrences. Voici, à travers trois exemples, une autre catégorie facilement repérable :

[13] l'agression d'un gilet rouge [agent SNCF de l'accueil en gare]

[14] Épreuve de force entre les robes noires et la Chancellerie

[15] Le Britannique Chris Boardman fut le premier maillot arc-en-ciel.

Il s'agit ici de l'habit caractéristique (source) pour l'individu qui le porte (cible).

Subjectivement, on peut être tenté de regrouper sabres, baguettes, gilets, robes et maillots. En effet, dans tous ces cas, la source est un objet en contact avec un individu et la cible un individu. Ceci nous mène à une taxinomie. L'activité classificatoire remonte à l'Antiquité, mais un classement indiscutable n'a jamais été atteint :

¹ Il s'agit là d'un lapsus mixte sémantique-formel.

« La rhétorique ancienne [a] tenté, indéfiniment, de dresser un relevé exhaustif des prétendues espèces de métonymie et de synecdoque, sans jamais parvenir à faire un partage parfait » (Henry 1971 : 18). Les auteurs modernes ont-ils eu plus de succès ? Beaucoup ont publié un inventaire partiel, reconnaissant l'existence d'un résidu d'inclassables ou la possibilité d'une variation infinie (v. p. ex. Bacry 1992). Parfois une pirouette permet de contourner l'obstacle, comme chez Ullmann (1969 : 285) pour qui « la classification traditionnelle des métonymies selon les divers types de contiguïté n'a qu'un modeste intérêt psychologique. » Il est rare qu'on trouve une typologie ambitieuse, comme celle de Morier (1961), qui décrit 28 catégories. La métonymie a bien été l'objet d'un regain d'intérêt avec le courant de la linguistique cognitive, mais là encore, on hésite devant l'entreprise classificatrice et ses obstacles : Riemer (2003) signale l'indétermination de nombreuses occurrences et ses hésitations quant à leur catégorisation. Radden (2003) évoque le fait que les métonymies, comme les espèces naturelles, ont des frontières floues et manifestent de la prototypie. Kövecses et Radden (1998) mentionnent les relations « indéterminées » (voir plus bas).

Ces derniers proposent une taxinomie élaborée. Les métonymies qui nous intéressent ici constituent pour eux la classe des « métonymies de concept », qui opèrent à un même niveau ontologique et impliquent un glissement d'un concept à l'autre. En cognitivistes, Kövecses et Radden proposent une typologie non pas des métonymies de discours (*linguistic metonymies*), mais des relations conceptuelles susceptibles de leur donner naissance. Ces relations forment deux super-classes : a) entre un domaine entier et ses parties et b) entre les parties d'un domaine (les auteurs ne parlent pas de domaine, mais d'*ICM* – j'y reviendrai). Si je les ai bien lus, la super-classe « domaine entier/parties » comprend six classes dont quatre réversibles avec deux catégories chacune, soit dix catégories en tout :

- le tout pour la partie et l'inverse² ;
- l'objet pour sa matière et l'inverse ;
- sous-événements successifs pour l'événement complexe ;
- sous-événements co-présents pour l'événement complexe ;
- la catégorie pour un de ses membres et l'inverse ;
- la catégorie pour une de ses propriétés et l'inverse.

Pour leur seconde super-classe, l'inventaire se complique, car le type de domaine constitue un niveau taxinomique ; je ne détaillerai pas les catégories, et le lecteur intéressé voudra bien se référer à l'article :

- domaine actionnel ;
- l'instrument pour l'action ;
- l'agent pour l'action ;

² Je ne traduis pas leurs intitulés avec une fidélité absolue.

et 10 autres ;

- domaine perceptif ;
- la perception pour la chose perçue et l'inverse ;
- domaine causatif, classe complexe avec plusieurs niveaux dont les plus hauts sont la cause pour l'effet et l'effet pour la cause ;
- domaine productif : 3 catégories ;
- domaine du contrôle : 2 ;
- domaine possessif : 2 ;
- domaine de la contention (conteneur et contenu) : 2.

En outre, des domaines de catégories différentes peuvent se combiner.

Un moyen d'évaluer une telle taxinomie est de la confronter à des occurrences. Pour ce faire, j'ai pris quelques exemples moins évidents que les sabres et les baguettes.

[16] Si ta besoin d'un truc je suis pas du tout une pointure en la matière mais je peux te filer qqcs conseils. [*sic*]

Il semble qu'on pourrait voir là une métonymie de la classe « un domaine et ses parties », et la catégorie « une propriété pour la catégorie », mais ceci est peu convaincant car il resterait à comprendre pourquoi la pointure d'un individu définit sa compétence et son renom³. De la même façon, il ne semble pas que les deux exemples suivants se laissent ranger facilement dans une des catégories de Kövecses et Radden :

[17] Les Verseau se lassent très vite de tout.

[18] Le mercure sera plus clément.

Je ne veux pas faire un mauvais procès aux auteurs, qui ne présentent pas explicitement leur taxinomie comme exhaustive. En outre, on pourrait toujours objecter que les exemples ci-dessus sont la trace en discours d'un empilement de métonymies plus simples. Il n'en reste pas moins qu'ils ne sont pas idiosyncrasiques ou liés à un contexte étroit, contrairement à l'exemple cité partout :

[19] Le sandwich au jambon réclame son addition

qui n'entre selon Kövecses et Radden dans aucune des catégories classiques et ne peut être interprété que dans le contexte d'un restaurant. Ces auteurs parlent de « domaine impliquant des relations indéterminées », notion qui semble elle-même

³ Sinon par une analogie quelque peu tirée par les cheveux entre la taille des pieds et l'importance du personnage, ce qui impliquerait une dose de métaphore. Noter que le TLF informatisé ne donne pas le sens métonymique correspondant.

quelque peu indéterminée... On peut remarquer en fait que cette métonymie est apparentée à l'exemple lui aussi présent dans la littérature :

[20] L'ulcère à l'estomac a encore bu du café

qui est, lui, interprétable dans un contexte hospitalier, et qu'en cherchant bien, on pourrait sans doute formuler une étiquette pour ce qui semble bien constituer une catégorie isolable, dont voici un autre exemple, fabriqué par moi :

[21] Le joint de culasse attend au café du coin

compréhensible dans un garage. Ces métonymies sont effectivement étroitement dépendantes du contexte, mais c'est parce qu'elles ne sont pas lexicalisées ; ce qui compte, c'est qu'il existe un patron réemployable par analogie.

Blank (1999) voit quant à lui deux super-classes de catégories de contiguïté bien différentes : la co-présence et la succession. Le classement des métonymies, on le voit, est encore loin d'être achevé et consensuel. Si les deux paires de super-classes de Kövecses et Radden et de Blank ne sont sans doute que deux abstractions différentes, il importe avant de pouvoir se placer à ce niveau de compléter l'inventaire des catégories de base, car il semble difficile de comprendre le phénomène sans savoir quoi faire d'un résidu d'inclassables.

4. LE NOM COMPOSÉ

Le nom composé [NN]_N présente des difficultés semblables. Je vais traiter ici principalement du composé français du type *timbre-poste*, mais son équivalent anglais manifeste les mêmes phénomènes sémantiques. Un composé *timbre-poste* est un composé endocentrique non-équatif-analogique (Arnaud, 2003). Une telle définition négative est révélatrice : il est impossible de trouver un terme simple qui subsume les liens de modification de N1 par N2.

La question du classement sémantique des relations internes aux noms composés est déjà présente, pour l'anglais, chez Maetzner (1860) et la littérature ultérieure fait état de taxinomies d'ampleur variable, dont un exemple très développé se trouve chez Adams (1973 : 57sq.), dans lequel je compte 23 catégories transposables aux composés *timbre-poste*. Comme pour les métonymies, on retrouve d'une taxinomie à l'autre des relations semblables et fréquentes, comme celle de matière constituante ou de but, mais il n'y a jamais recouvrement complet, et les taxinomies englobent fréquemment des catégories établies sur des principes disparates. De nombreux auteurs, comme Jespersen (1942/1961 : 142) présentent leur classement comme non exhaustif, ce dernier qualifiant de « *endless* » l'inventaire des relations possibles, comme Barbaud (1971) après lui sur le français, qui écrit : « Nous n'en finirions pas de relever les diverses complémentarités possibles à l'intérieur de la structure binominale ». Enfin, Maetzner (1860) évoquait déjà la difficulté de décrire la nature précise de certaines relations, constatation qu'on retrouve un siècle plus tard chez Marchand (1960 : 22).

Lorsqu'on cherche à classer sémantiquement une collection étendue de composés, on se trouve vite confronté à des problèmes similaires à ceux déjà évoqués. De façon prévisible, des relations saillantes et productives apparaissent d'emblée, du genre de « N2 est une des parties de N1 », qu'on trouve par exemple dans *tampon-buvard* ou *tailleur pantalon*, mais on rencontre des unités qui ne se laissent pas classer si facilement, et cela à cause d'une apparente indétermination. Contrairement aux deux exemples précédents, certains des liens sont difficiles à saisir, presque ineffables sans une abstraction considérable ; ainsi, la relation que je pense être présente dans *cas-sujet*, *année-lumière*, et qu'en désespoir de cause j'ai étiquetée « N2 'a' N1 ». La deuxième raison principale est qu'on se trouve très hésitant devant d'autres unités, qui un instant semblent pouvoir être rattachées à une relation établie, et l'instant d'après basculent dans une autre interprétation. La solution est d'affecter plusieurs étiquettes à ces unités, telle *base-vie*, qui relève à la fois de « N2 est la destination, l'objet de N1 » et de « N2 a lieu dans N1 ».

Une telle entreprise pose évidemment la question de la granularité désirable, et par voie de conséquence celle de regroupements et donc, de nouveau, d'une taxinomie. Mon traitement de 800 composés *timbre-poste* a abouti, non sans palinodies, à la formulation de pas moins de 54 catégories de base, ce qui reflète la variété de ces composés en regard de l'unicité des équatifs-analogiques, qui rappelle à son tour la variété de la métonymie face au monolithisme analogique de la métaphore. Comme pour la métonymie, des regroupements apparaissent d'eux-mêmes, plusieurs relations relevant par exemple d'une relation d'inclusion, comme « N2, concret et discret, est l'une des parties de N1 » (p. ex. *fourgon-pompe*), « N2, concret et continu, est entièrement constituant de N1 » (ex. *semelle-crêpe*), et les 54 catégories de base se sont laissées regrouper en huit super-catégories. Il est probable que l'application de cette taxinomie à des composés anglais rendrait nécessaire la formulation de catégories de base supplémentaires.

5. PROCESSUS

Les phénomènes que nous avons envisagés jusqu'ici relèvent tous d'une sous-spécification sémantique de discours, dont il faut chercher à expliquer les mécanismes.

L'explication devra tenir compte du fait que la forme présente en discours est issue de processus, lesquels opèrent sur des représentations conceptuelles et/ou lexicales. Dans le cas des lapsus, les modèles psycholinguistiques (voir Arnaud 1999) font généralement appel à la notion d'activation. Les détails varient, mais le processus communément admis est que lors de la phase d'idéation, le concept-cible est activé, mais l'activation atteint également les concepts sémantiquement voisins. Si le niveau d'activation d'un de ceux-ci dépasse celui de la cible pour une raison ou une autre (fluctuations aléatoires, activation rémanente, etc.), c'est lui qui envoie de l'activation vers les couches suivantes, notamment celle des représentations phonologiques.

Depuis Aristote, la métonymie est censée reposer sur la contiguïté. Ce terme, qui n'engage à rien sur le processus, a longtemps été accepté sans discussion, et ce n'est qu'avec l'éclairage cognitiviste qu'on a cessé d'y voir une contiguïté référentielle pour penser en termes de contiguïté conceptuelle. La contiguïté trouve une autre formulation avec Lakoff et Johnson (1980), et le processus est celui de l'application (*mapping*) d'un ensemble vers l'autre. Pour la métaphore, l'application a lieu entre deux *ICM* (*Idealized Cognitive Models*), entités sur lesquelles je reviendrai ci-dessous, alors que la métonymie implique une application à l'intérieur d'un même *ICM*. Un autre terme, dû à Langacker, *domain highlighting*, « mise en relief de domaine » est également employé (Croft 1993), sans qu'il semble y avoir de différence fondamentale dans le processus postulé. La notion centrale d'application est généralement acceptée, mais on en reste souvent là, ce qui au fond ne nous amène guère plus loin que la contiguïté et ne nous dit pas *pourquoi* il y a des métonymies. Kövecses et Radden (1998) vont plus loin et exposent les principes cognitifs qui semblent sous-tendre les métonymies conceptuelles. Si la source est utilisée pour référer à la cible, c'est qu'elle est dotée d'une saillance élevée et constitue de ce fait un point d'accès facile. Trois principes déterminent la saillance relative : l'expérience humaine (l'humain a priorité sur le non-humain, le concret sur l'abstrait, etc.), la sélection perceptive (certains stimuli ont priorité), et les préférences culturelles. À ces principes s'ajoutent des principes communicatifs. Plus grand sera le nombre de principes qui s'appliqueront, plus forte sera la motivation d'une métonymie, à moins que d'autres motivations, stylistiques par exemple, ne prennent le dessus.

Si elles sont majoritaires, les vues ci-dessus ne sont pas les seules, et il existe une approche plus syntaxique de la métonymie. Warren (1999) y voit ainsi le résultat d'un processus où une tête virtuelle est effacée au profit de son modifieur, comme dans :

[24] *The kettle is boiling*

qui proviendrait de

[that [which is in the kettle]] is boiling.

tête modifieur

On retiendra toutefois que, quel que soit le niveau, conceptuel ou linguistique, où on place le processus métonymique qui établit les liens source-cible, ce sont les représentations qui permettent l'existence de ceux-ci.

Le nom composé invitait par sa compacité les premiers linguistes qui se sont intéressés à lui à évoquer les prépositions implicites ou un processus syntaxique venant réduire une proposition, mais les modèles les plus achevés sont bien sûr dus à la Grammaire générative, avec les travaux pionniers de Lees (1960). Ces modèles postulent l'existence en structure profonde d'une phrase dont la prédication matérialise la relation sémantique N1-N2, et qui est reprise par une série de transformations pour aboutir au composé. L'absence d'intérêt de la théorie générative pour le lexique ne menait pas à expliquer d'où sortait le verbe effacé, question qui aurait pourtant dû paraître cruciale. L'abandon des transformations allait sonner le glas de

cette approche des composés, et depuis Lieber (1983), il semble que la modélisation dans un cadre syntaxique des processus de composition nominale n'excite plus les chercheurs. C'est du côté des représentations qu'il convient de rechercher la nouveauté.

6. REPRÉSENTATIONS

Les processus ne peuvent opérer que sur des représentations avec lesquelles ils sont compatibles. La contiguïté impliquée dans la métonymie a été classiquement comprise comme une co-présence dans le même champ lexical (Henry 1971 : 26). On a vu plus haut les critiques dont a été l'objet la notion de champ, notamment celles qui ont porté sur l'impossibilité de trouver des critères objectifs de délimitation, et Blank (1999) fait remarquer en outre que dénommer *contiguïté* le fait que des unités soient dans des champs que nous avons créés *ad hoc* pour rendre compte de l'*impression* de contiguïté revient à confondre notre conceptualisation des choses et leur nature.

La linguistique cognitive a apporté des raffinements considérables à la notion de champ. Comme on l'a vu, Lakoff (v. Lakoff et Johnson 1980) est à l'origine de la notion d'*ICM* (*Idealized Cognitive Model*), qui ne semble pas être fondamentalement différente de celle de *domaine* chez Langacker (1987 : 147-182). Dans la théorie de Langacker, les concepts incluent toutes les informations liées à un *denotatum*, et donc ne possèdent sans doute pas de frontière nette, ces informations pouvant varier avec le temps, être imprécises, etc. Les concepts ont un cœur, qui comprend les informations obligatoirement activées à chaque accès, et une périphérie, dont l'activation des éléments est seulement facilitée. Comprendre le sens d'un mot, c'est accéder au réseau d'informations correspondant par un ou plusieurs points d'entrée et activer le réseau (Croft 1993). Les domaines sont le résultat émergent de l'organisation de l'information sémantique en fonction de l'expérience. À l'intérieur d'un domaine, l'ensemble d'informations constituant un concept se détache sur le fond comme une *Gestalt*. Les domaines peuvent s'emboîter ou être en intersection, constituant des *matrices de domaines* à l'intérieur desquelles la métonymie peut opérer par application d'un concept vers l'autre.

Les critiques qui ont été formulées contre les domaines, notamment par Croft (1993), Warren (1999), Riemer (2003), ne sont pas fondamentalement différentes de celles qui ont porté sur les champs lexicaux ou sémantiques : la notion de domaine reste trop floue par manque de critères implémentables, et la délimitation d'un domaine pour rendre compte d'un lien sémantique est souvent *ad hoc*.

Les modèles psycholinguistiques utilisés pour rendre compte des lapsus sont globalement des modèles à information décomposée et diffuse. L'activation correspondant à un concept s'applique, à des niveaux divers, à un ensemble d'éléments. Certains de ces éléments font également partie d'autres ensembles correspondant à d'autres concepts. Les détails varient, comme la présence ou l'absence de nœuds conceptuels, mais ce que nous retiendrons est cette extrême dispersion. Ceci cor-

respond aux modèles connexionnistes, qui se prêtent à des simulations (Harley 1993).

La recherche sur les noms composés à complémentation n'a pas mené à la formulation de modèles de représentations, mais leur proximité de la métonymie incite à favoriser les modèles diffus et en réseaux, seuls capables de rendre compte des relations « inclassables ». Il convient toutefois de citer une recherche qui s'est orientée dans une autre direction, puisqu'elle s'est inspirée du *Generative Lexicon* de Pustejovsky (1995). Bassac et Bouillon (2001) ont tenté d'expliquer le fonctionnement des composés téléliques du turc et Bassac (2003) montre comment le modèle de représentation des unités lexicales pouvait rendre compte des différences entre quatre types de composés de cette langue, pour lesquels des *qualia* différents sont saturés. Les représentations pustejovskiennes sont certes riches en information, mais elles ne peuvent prétendre à la souplesse des modèles en réseau. Il reste à voir si elles peuvent rendre compte de la variété des relations observées dans les composés *timbre-poste* ou dans les composés [NN]_N anglais.

7. CONTEXTE

La prise en compte du contexte ne peut en principe pas avoir lieu dans les représentations. Certes, celles-ci ne sont pas immuables, mais elles ne peuvent guère à la fois manifester de la permanence et changer au gré des circonstances. C'est donc plutôt du côté des processus qu'il faut chercher. Pour les lapsus, les choses ne sont pas trop difficiles à modéliser : tout élément du contexte (perception de choses ou de discours, pensée obsédante, etc.) ou du cotexte (persévération, anticipation) est susceptible d'activer les représentations correspondantes, lesquelles peuvent finir par avoir une activation plus élevée que celle de la cible, d'où les lapsus contaminationnels ou mixtes.

Métonymie et noms composés ont en commun de pouvoir être lexicalisés ou occasionnels. Pour ce qui est de la métonymie, nous avons vu qu'une métonymie occasionnelle et hautement contextuelle comme celle du sandwich au jambon pouvait sans doute être rattachée à une catégorie. Reste à savoir si l'on trouve des catégories véritablement inédites. Quant aux noms composés, s'ils sont lexicaux, ils pré-existent au contexte. Seules les suites NN ou AdjN occasionnelles nécessitent la pré-activation de N1 et de N2 et un calcul du sens, comme pour le *blind man* de l'histoire grivoise *The Nuns and the Blind Man* que je laisserai au lecteur le soin d'imaginer, qui s'avère être, non pas un aveugle, mais un poseur de persiennes.

BIBLIOGRAPHIE

- Adams (V.), 1973 : *An Introduction to Modern English Word-Formation*, London, Longman.
 Arnaud (P.), 1999 : « Target-error resemblance in French word substitution speech errors and the mental lexicon », dans *Applied Psycholinguistics*, 20, pp. 269-87.
 Arnaud (P.), 2003 : *Les Composés timbre-poste*, Lyon, Presses Universitaires de Lyon.

- Bacry (P.), 1992 : *Les Figures de style*, Paris, Belin.
- Barbaud (P.), 1971 : « L'ambiguïté structurale du composé binominal », dans *Cahiers de Linguistique* (Montréal), 1, pp. 71-116.
- Blank (A.), 1999 : « Co-presence and succession : a cognitive typology of metonymy », dans Panther (K.-U.) et Radden (G.), édit., *Metonymy in Language and Thought*, Amsterdam : Benjamins, pp. 169-191.
- Croft (W.), 1993 : « The role of domains in the interpretation of metaphors and metonymies » dans *Cognitive Linguistics*, 4, pp. 335-370.
- Garrett (M. F.), 1980 : « Levels of processing in sentence production » dans Butterworth (B.), édit., *Language Production, Vol. I : Speech and Talk*, London, Academic Press, pp. 177-220.
- Harley (T. A.), 1993 : « Phonological activation of semantic competitors during lexical access in speech production » dans *Language and Cognitive Processes*, 8, pp. 291-309.
- Henry (A.), 1971 : *Métonymie et Métaphore*, Paris, Klincksieck.
- Hotopf (W. H. N.), 1980 : « Semantic similarity as a factor in whole-word slips of the tongue » dans Fromkin (V. A.), édit., *Errors in Linguistic Performance*, New York, Academic Press, pp. 97-109.
- Hotopf (W. H. N.), 1983 : « Lexical slips of the pen and tongue » dans Butterworth (B.), édit., *Language Production, Vol. 2*, London, Academic Press, pp. 147-199.
- Jespersen (O.), 1961 : *A Modern English Grammar on Historical Principles*, London, Allen & Unwin [1^{re} édit. 1942].
- Kövecses (Z.) et Radden (G.), 1998 : « Metonymy : developing a cognitive linguistic view » dans *Cognitive Linguistics*, 9, pp. 37-77.
- Lakoff (G.) et Johnson (M.), 1980 : *Metaphors we Live by*, Chicago, University of Chicago Press.
- Langacker (R. W.), 1987 : *Foundations of Cognitive Grammar. Vol. I : Theoretical prerequisites*, Stanford, Stanford University Press.
- Langacker (R. W.), 1993 : « Reference-point constructions » dans *Cognitive Linguistics*, 4, pp. 1-38.
- Lees (R. H.), 1960 : *The Grammar of English Nominalizations*, Bloomington (Indiana), Indiana University.
- Lehrer (A.), 1974 : *Semantic Fields and Lexical Structure*, Amsterdam, North-Holland.
- Lieber (R.) 1983 : « Argument linking and compounds in English » dans *Linguistic Inquiry*, 14, pp. 251-285.
- Maetzner (E.), 1873 : *Englische Grammatik*, Weidmann [1^{re} édit. : 1860].
- Marchand (H.), 1969 : *The Categories and Types of Present-day English Word-formation*, München, Beck.
- Morier (H.), 1961 : *Dictionnaire de poétique et de rhétorique*, Paris, PUF [édit. 1989].
- Pustejovsky (J.), 1995 : *The Generative Lexicon*, Cambridge (Massachusetts), M.I.T. Press.
- Radden (G.), 2003 : « How metonymic are metaphors ? » dans Dirven (R.) et Pörings (R.), édit., 2003 : *Metaphor and Metonymy in Comparison and Contrast*, Berlin, Mouton-de Gruyter, pp. 407-434.
- Riemer (N.), 2003 : « When is a Metonymy no longer a Metonymy ? » dans Dirven (R.) et Pörings (R.), édit., 2003 : *Metaphor and Metonymy in Comparison and Contrast*, Berlin, Mouton-de Gruyter, pp. 379-406.
- Ullmann (S.), 1969 : *Précis de sémantique française*, Berne, Francke [4^e édit.].
- Warren (B.), 1999 : « Aspects of referential metonymy » dans Panther (K.-U.) et Radden (G.), édit., *Metonymy in Language and Thought*, Amsterdam, Benjamins, pp. 121-135.
- Wierzbicka (A.), 1985 : *Lexicography and Conceptual Analysis*, Ann Arbor, Karoma.

INFÉRENCES ET RÔLE DU MICROCONTEXTE ET DU MACROCONTEXTE

Anna Anastassiadis-Syméonidis et Olympia Tsaknaki
département de philologie, section de linguistique
Université Aristote de Thessaloniki
Thessaloniki, Grèce

Résumé : Pour arriver à interpréter un texte par inférence l'auditeur/lecteur n'a à sa disposition que le contexte linguistique et extralinguistique ainsi que les informations livrées par le registre de langue. Nous examinons le facteur contexte et ses multiples faces que l'auditeur/lecteur doit prendre en compte pour aboutir à l'interprétation métaphorique. Nous prenons comme exemples-pivots les adjectifs dénominaux du grec moderne en – inos. Dans le but de vérifier les règles d'inférences proposées, nous présentons les résultats de notre recherche sur un corpus ¹.

Mots-clés : macrocontexte, noms de matériau, inférences.

1. INTRODUCTION

Il est bien connu que c'est le contexte qui permet d'attribuer à une forme une interprétation plutôt figurée que littérale (*i.e.* Marque-Pucheu 2001 : 16). Il s'agit du contexte linguistique et extralinguistique (connaissance du monde) ainsi que du macrocontexte et du microcontexte. Pour le cas du macrocontexte, c'est-à-dire dans les relations interlexicales (entre le mot et son entourage), nous prenons comme exemples-pivots les adjectifs dénominaux du grec moderne en – *inos* dont le nom de base est un nom de matériau ² : ces adjectifs construits peuvent avoir, suivant le nom recteur, une lecture/interprétation extensionnelle, ex. *petrino spiti* « maison en pierre » ³ ou intensionnelle *petrini kardhia* « cœur de pierre ». De même, pour le cas du microcontexte, c'est-à-dire dans les relations intralexicales (entre les morphèmes d'un mot construit), un exemple-pivot serait les substantifs dérivés du grec moderne en – *ia* construits sur une base nominale désignant soit une partie du corps, ex. *angonas* « coude » < *angonia* « coup de coude », soit un fruit, ex. *lemoni* « citron » < *lemonia* « citronnier ». Ce travail porte sur le cas du macrocontexte.

¹ Nous tenons à remercier Kyriaki Ioannidou, étudiante de DEA, Université Aristote de Thessaloniki, Ionas Michailidis, doctorant, Université de Tours et Gérard Petit, maître de conférences, Paris X.

² Il s'agit de substantifs répondant à la question : « en quoi est fait X ? ».

³ Traduction.

2. CADRE THÉORIQUE DE LA MÉTAPHORE

En ce qui concerne la métaphore, elle relève de la sémantique et de la syntaxe, car, d'une part, on parle généralement de métaphore en cas de transgression de la distribution sémantique, c.-à-d. des règles de sélection, des critères d'incompatibilité de champs sémantiques et d'inconvenance référentielle à l'intérieur d'un champ sémantique, et, de l'autre, de la distribution syntaxique : *η καρδιά του είναι/έγινε πέτρα* [son cœur est/est devenu pierre]⁴ « s'est endurci ». Nous avons affaire à une métaphore nominale et il faut chercher le trait servant de pivot métaphorique. Le N1 *καρδιά* « cœur » à fonction référentielle désignative est en position de thème. Le N2 *πέτρα* « pierre » à fonction dénotative de propriétés est en position de prédicat. L'absence de déterminant devant le N2 indique son statut non désignatif et sa dépendance référentielle de N1 (Salteri-Cacouros et Tamba 2001 : 62). Par conséquent, le terme métaphorique n'acquiert son interprétation que dans une relation syntaxique entre termes, entre sujet et prédicat le cas échéant (Marque-Pucheu 2001 : 10). La métaphore s'inscrit donc dans la configuration N est (un) N (Tamine 1979 : 66-67), localisation syntaxique de la recatégorisation, cf. aussi la construction NN *μια καρδιά πέτρα* [un cœur pierre] (Noailly 1996).

La métaphore intervient dans le discours mais aussi dans la formation lexicale, nous plaçant alors dans la langue (Marque-Pucheu 2001 : 18) : *Πέτρα* dans sa lecture intensionnelle est une création consignée en langue, puisqu'il y a un consensus là-dessus entre sujets. Par ailleurs, il s'agit d'une extension de sens qui est prédictible de notre point de vue, comme nous allons voir.

Pour Kleiber (1994 : 43), qui évite de recourir au sens littéral, la métaphore est une infraction à une catégorisation GN être GN. La tendance dominante actuellement est d'évoquer un sens figuré en cas de catégorisation induite (Kleiber 1999 : 133) : l'analogie, fondamentale dans l'interprétation des énoncés, ainsi que dans leur accès à l'interprétabilité, va de pair avec l'hétérogénéité des catégories assimilées. L'analogie s'établit selon un minimum de traits partagés. Selon Prandi (1992 : 29), qui place les énoncés figurés à mi-chemin entre énoncés acceptables et énoncés contradictoires, la figure provoque un conflit entre concepts. Pour Prandi (1992 : 80), la métaphore défie l'identité conceptuelle du référent.

Enfin, pour Salteri-Cacouros (2001 : 58), un thème (GN1) a un ancrage référentiel autonome et une portée qui s'étend jusqu'à la clôture de l'énoncé. Il fixe donc, une fois pour toutes, l'identité du référent auquel s'applique l'ensemble du prédicat. Ce référent pourra alors être le support d'une métaphorisation dont il fournira le thème métaphorisé, sans pour autant perdre son identité. Ainsi sera établie l'orientation référentielle de cette structure, qui commande l'annulation des traits sémantiques non pertinents du N2. Car le N2 en tant que noyau du prédicat nominal ne renvoie pas à un référent situationnel individué, mais à un ensemble de représentations inférables à partir de la notion qu'il dénomme. C'est pourquoi la caractéristique sélectionnée est le plus souvent spécifiée dans le contexte immédiat (p. 59). La fixation référentielle est le facteur qui conditionne l'orientation du jugement méta-

⁴ Traduction littérale.

phorique. Le sujet est comparé au prédicat (p. 62). L'interprétation métaphorique repose crucialement sur l'identification précise d'un référent désigné par le GN1 et sur la représentation qu'en donne son assimilation à une autre notion. Il y a instauration d'une relation analogique en catégorisant le GN1 métaphorisé au moyen d'un GN2 métaphorisant. Le processus général d'interprétation se fonde sur un jugement analogique. Un tel processus repose sur l'assimilation d'un référent bien identifié à une image mentale de quelque autre notion.

Par conséquent, la métaphore, en accord avec son sens étymologique, serait, du point de vue sémantique, un transfert, par analogie, ressemblance, comparaison ou similitude (*i.e.* Prandi 1992 : 220), entre différents domaines référentiels ; il y aurait une incompatibilité sémique, ce qui entraîne une nouvelle distribution syntaxique.

L'analyse que nous proposons tend à unifier sens littéral et sens figuré, chaque interprétation représentant une des faces d'un polyèdre régulier ou plutôt une zone sur la surface d'une sphère. L'ensemble des interprétations d'une unité lexicale constituerait son sens (*i.e.* Anastassiadis-Syméonidis 2004 : 30). Ce point de vue va dans le même sens que la notion de motifs comme principes unificateurs de la diversité lexicologique (Cadiot et Visetti 2001 : 102).

3. CORPUS UTILISÉ

L'analyse fondée sur des *corpus* présente beaucoup d'avantages, car elle minimise la subjectivité et met en valeur des phénomènes moins accessibles ou impossibles à traiter par intuition. Si le *corpus* est représentatif, les données soumises à l'analyse sont à la fois authentiques et attestées et avant tout objectives.

L'étude attentive de notre *corpus* nous a permis de trouver des exemples concrets qui confirment la proposition présentée ci-dessus. Nos observations ont été établies à partir d'un *corpus* constitué des articles du journal grec *TA NEA* couvrant la période de l'année 2000 disponible sous forme électronique. La longueur totale du *corpus* consulté est environ de 5 000 000 mots. Notre recherche sur ce *corpus* de textes journalistiques nous a permis de faire un relevé systématique des adjectifs en – *inos*. Le *corpus* contenait aussi des emplois non métaphoriques de cette catégorie d'adjectifs, qui ont été exclus après une analyse sémantique. Les résultats ont aussi révélé des adjectifs substantivés et des adverbes, qui ont été exclus de notre *corpus*.

Après une analyse de nos données, nous avons eu la possibilité d'identifier les adjectifs en – *inos* dont l'emploi serait considéré comme métaphorique. Nous avons examiné leur contexte et nous avons focalisé notre attention sur leur sens et emploi.

Le repérage des métaphores est fait à l'aide des concordances, c'est-à-dire des mots de la requête accompagnés du contexte. Les emplois qui ne rentraient pas dans le cadre de notre recherche ont été retranchés manuellement.

4. CONTEXTE LINGUISTIQUE ET EXTRALINGUISTIQUE : LE SUFFIXE –*INOS*

Il ne faut pas négliger l'aspect sémantique et productif de toute une série de régularités de noms de matériau qui, en développant un sens intensionnel, fonctionnent comme des noms de qualité, pouvant servir, de cette façon, de noms de base d'adjectifs dénominaux de qualité. Ces adjectifs dérivés sont construits au moyen des suffixes – *in* (*os*) et – *eni* (*os*), qui servent par excellence à construire des adjectifs dont la base est un nom de matériau du vocabulaire général (Anastassiadis-Syméonidis 1997). Dans notre *corpus*, nous ne nous sommes intéressées qu'aux adjectifs en – *inos*. Par la systématisme qu'implique le transfert d'un domaine à l'autre, ces éléments lexicaux sont alors directement interprétables, comme nous pouvons constater à l'aide des adjectifs que nous avons rencontrés dans notre *corpus* : *αδάμας* « diamant » → *αδαμάντινος*, *αέρας* « air » → *αέρινος*, *αίμα* « sang » → *αιμάτινος*, *αλάβαστρος* « albâtre » → *αλαβάστρινος*, *ατσάλι* « acier » → *ατσάλινος*, *βελούδο* « velours » → *βελουδίνος*, *γη* « terre » → *γήινος*, *γυαλί* « verre » → *γυάλινος*, *κρύσταλλο* « cristal » → *κρυστάλλινος*, *λίθος* « pierre » → *λίθινος*, *μπρούτζος* « bronze » → *μπρούτζινος*, *ξύλο* « bois » → *ξύλινος*, *πέτρα* « pierre » → *πέτρινος*, *πετσι* « peau » → *πέτσινος*, *πηλός* « terre cuite » → *πήλινος*, *πυρ* « feu » → *πύρινος*, *σάρκα* « chair » → *σάρκινος*, *ύδωρ* « eau » → *υδάτινος*, *χαλκός* « bronze » → *χάλκινος*, *χάλυβας* « acier » → *χαλύβδινος*, *χαρτί* « papier » → *χάρτινος*.

Nous avons aussi répertorié des adjectifs de couleur comme *πράσινος* « vert », *κόκκινος* « rouge », *κίτρινος* « jaune », *ρόδιος* « rose » et plus rarement *πρασινοκίτρινος* « vert et jaune », *φαιοκόκκινος*⁵ « gris et rouge » etc. Ces adjectifs dérivés sont formés sur une base qui est un substantif et le suffixe –*inos* formateur d'adjectifs : *ρόδιος* < *ρόδο* « rose » + –*inos*. Quelques adjectifs sont considérés comme des signes linguistiques non motivés aujourd'hui : *πράσινος* < *πράσο* « poireau » + –*inos*, *κόκκινος* < *κόκκος* « grain tinctorial » + –*inos*, *κίτρινος* < *κίτρο* « cédrat » + –*inos* (Anastassiadis-Syméonidis 1987 : 401).

Mais les variations de sens ne sont pas uniquement déterminées par le contexte linguistique. On les déduit aussi du contexte extralinguistique. L'interprétation s'appuie en général sur des éléments saillants du contexte et sur des représentations comportant des savoirs encyclopédiques concernant les référents mis en rapport lors du processus métaphorique. Prenons la phrase *Το σπίτι είναι πέτρινο* « La maison est en pierre » et l'exemple suivant qui figure dans le *corpus* :

[1] Θα όφειλα να έχω σιδερένιο μέτωπο και πέτρινη καρδιά

« J'aurais dû avoir un front de fer et un cœur de pierre »

⁵ Le trait [± savant] du grec moderne concerne aussi le domaine des adjectifs de couleur. Le grec moderne dispose de paires d'adjectifs désignant la même couleur dont l'un est savant et l'autre non savant : *φαιός/γκρι* « gris », *ερυθρός/κόκκινος* « rouge » (Anastassiadis-Syméonidis 1987 : 394). L'adjectif *φαιοκόκκινος* est un adjectif composé qui provient des adjectifs : [+ savant] *φαιός* + [+/- savant] *κόκκινος*.

Le mot *πέτρα* « pierre »⁶ n'est pas employé de manière arbitraire par rapport soit à son sens habituel soit à son sens métaphorique résultant de l'association de *πέτρα* « pierre » et *καρδιά* « cœur ». Pareillement pour *κράτος* « état », *στάση* « attitude » et *χρόνια* « années » dans les cas de *πέτρινο Κράτος* « État de pierre », *πέτρινη στάση* « attitude de pierre », *πέτρινα χρόνια* « années de pierre ». La pierre étant, selon le stéréotype, un matériau très dur, le contexte implique que le cœur, l'État, l'attitude et les années sont très durs.

L'analyse présentée ci-dessus s'applique aux métaphores conventionnelles (Sperber & Wilson 1989 : 353) (*ξύλινος λόγος* « discours de bois », *ξύλινη γλώσσα* « langue de bois », *κρυστάλλινος ήχος* « son cristallin », *βελούδινη φωνή* « voix de velours », *πύρινος λόγος* « discours enflammé », *ασάλινα νεύρα* « nerfs d'acier », *κίτρινος Τύπος* [presse jaune] « presse à scandale », *κόκκινη Πρωτομαγιά* « Premier mai rouge »), mais nous estimons qu'elle peut aussi être appliquée aux métaphores vives. Selon les données de notre *corpus*, les exemples suivants où un adjectif en – *inos* est colloqué avec un nom peuvent illustrer notre point de vue : *ξύλινο στερέοτυπο* « stéréotype de bois », *ξύλινη διγλωσσία* « double langage de bois », *κρυστάλλινη Ελληνίδα* « Grecque cristalline », *κρυστάλλινη συνέντευξη* « entretien cristallin », *πύρινος απολογισμός* « bilan enflammé », *πέτσινο ποδόσφαιρο* [football de cuir] « football sans valeur », *πέτρινοι κοιλιακοί* « abdominaux de pierre », *κόκκινες λίστες* « listes rouges », *Πράσινος υπουργός* « ministre Vert ».

Les adjectifs en – *inos* ont une lecture intensionnelle non seulement dans le cadre de la structure Adjectif + Nom mais aussi quand ils sont utilisés dans une structure prédicative :

[2] Τα προβλήματα δεν είναι ούτε κόκκινα ούτε πράσινα ούτε γαλάζια

[Les problèmes ne sont ni rouges ni verts ni bleus]

[3] Η θέση της κυβέρνησης είναι κρυστάλλινη και δεδομένη

[La position du gouvernement est cristalline et donnée]

[4] Δείτε πόσο βελούδινη και τρυφερή είναι αυτή η ταινία

[Regardez, combien ce film est tendre et velouté]

[5] Υπάρχουν δύο πιθανά σενάρια και κανένα δεν είναι ρόδινο

[Il y a deux scénarios probables et aucun n'est rose]

Les adjectifs en – *inos* peuvent aussi avoir une lecture intensionnelle dans des structures comparatives :

[6] Είναι πιο « ανθρώπινος » από την πραγματικότητα της πολιτικής

[Il est plus « humain » que la réalité de la politique]

[7] Είναι μία από τις πιο « χάρτινες » περιπέτειες από συστάσεως Χόλιγουντ

⁶ Il est à noter qu'à chaque étape diachronique, le sens littéral est accompagné d'un sens figuré/métaphorique différent : *πέτρα* « roc/roche » est à la base de la création du sens de « solide ». Par contre, *πέτρα* « pierre » est à la base de la création du sens de « dur » (v. Anastassiadis-Syméonidis 2004 : 31).

[Elle est une aventure de papier depuis la constitution d'Hollywood] ⁷

[8] Φρόντισε να κάνει όσο πιο... κόκκινες δηλώσεις μπορούσε

[Il a essayé de faire des déclarations les plus... rouges autant que possible]

Il est à noter que les adjectifs en – *inos* n'ont pas qu'un seul emploi. Leur sens varie en fonction du nom recteur. À titre d'exemple, nous présentons ci-dessous une liste d'occurrences concernant les divers emplois métaphoriques :

– γυαλί « verre » → γυάλινος «de verre »

[9] ONE: οικονομική επιτυχία ή οικονομία με γυάλινα πόδια; (fragilité)

[UEM : succès économique ou économie aux pieds de verre ?]

[10] Κοιτάει συνήθως με γυάλινο μάτι είτε το άπειρον είτε τις μύτες των παπουτσιών (inexpressivité)

[D'habitude, il regarde avec un œil de verre soit l'infini soit la pointe des pieds]

– κρύσταλλο «cristal » → κρυστάλλινος «de cristal »

[11] Έθεσε όλα τα θέματα που συζητήθηκαν με κρυστάλλινη διαύγεια (clarté)

[Il a posé toutes les questions qui ont été discutées avec une clarté cristalline]

[12] Ο ζωοδότης Ήλιος από πάνω προσφέρει ζεστές κρυστάλλινες δέσμες (lucidité)

[Le soleil nourricier au-dessus offre des faisceaux chauds en cristal]

[13] Τα νερά ήταν ορμητικά, αλλά κρυστάλλινα (pureté)

[Les eaux étaient impétueuses mais cristallines]

– χαρτί «papier » → χάρτινος «de papier »

[14] Οι λάτρεις της χάρτινης περιπέτειας θα το εκτιμήσουν (fausse aventure)

[Les amateurs de l'aventure de papier l'apprécieront]

[15] Στο «σκαμνί» για τη χάρτινη πολυκατοικία (fragile en parlant de construction)

[Sur la sellette pour l'immeuble de papier] ⁸

La connaissance encyclopédique rend certaines hypothèses hautement accessibles ; si celles-ci conduisent à des interprétations cohérentes avec le principe de pertinence, toutes les autres interprétations sont bloquées. L'adjectif de couleur est en lien direct avec le contexte, car une couleur n'a pas un seul et unique symbolisme, mais elle peut avoir des signifiés variables. Ainsi, quand un adjectif de couleur est inséré dans une phrase, il prend un sens qui provient du contexte. Si nous plaçons, à titre d'exemple, les adjectifs de couleur πράσινος « vert », κόκκινος « rouge » et κίτρινος « jaune » dans divers contextes, nous voyons qu'ils font appel à des significations différentes selon le contexte :

– πράσινος « vert » :

⁷ « Elle est une aventure des plus fausses depuis la constitution de Hollywood ».

⁸ « Sur la sellette pour le château de cartes ».

[16] Το... *πράσινο πρωτάθλημά του το πανηγύρισε έξαλλα* (= le vert est la couleur de l'équipe grecque Panathinaikos)

[Son championnat... vert, il l'a pavosé d'une manière délirante]

[17] Η Ν.Δ.⁹ *βάζει πόδι στον μέχρι σήμερα πράσινο αυτό νομό* (= le vert est la couleur du parti socialiste grec)

[La présence de la N. D. commence à se faire sentir à ce département qui était vert jusqu'aujourd'hui].

[18] *Είχε χαρακτηριστεί κίτρινη και μετά πράσινη κατά τον σεισμό της 7ης Σεπτεμβρίου* (= la couleur caractérise le degré de dangerosité des immeubles après un tremblement de terre : vert = sans risque, jaune = risque moyen, rouge = risque extrême)

[Il a été caractérisé jaune et par la suite vert pendant le tremblement de terre du 7 septembre].

[19] *Μια νέα πράσινη... τεχνολογία κινεί τον κόσμο* (= le vert symbolise le produit non polluant du point de vue écologique)

[Une nouvelle technologie... verte meut le monde »]

– κόκκινος «rouge » :

[20] *Κόκκινο Σαββατοκύριακο σε μπάσκετ, ποδόσφαιρο* (= le rouge est la couleur de l'équipe grecque Olympiakos)

[Week-end rouge au basket, football]

[21] *Οι βόρειοι, το «κόκκινο Βιετνάμ», τους αποθαρρύνουν με διάφορους τρόπους* (= le rouge symbolise l'idéologie/le parti communiste)

[Les Nordistes, le « Vietnam rouge », les découragent par différents moyens]

[22] *Τα κτήρια, λόγω ζημιών, έχουν χαρακτηριστεί κόκκινα ή κίτρινα* (= le rouge caractérise le degré de dangerosité après un tremblement de terre)

[Les immeubles, à cause de dégâts, ont été caractérisés rouges ou jaunes]

[23] *Κόκκινοι βαμμένοι δρόμοι* (= le rouge symbolise le sang)

[Des routes peintes en rouge]

– κίτρινος «jaune » :

[24] *Ο Τύπος οργιάζε, κυρίως ο λεγόμενος κίτρινος ή λαϊκός (το ίδιο κάνει)* (κίτρινος Τύπος «presse jaune » = presse à scandale)

[La presse s'emballait, notamment la presse dite jaune ou populaire (c'est la même chose)]

⁹ N. D. = *Nea Dimokratia* est un parti politique grec.

[25] Οι διαπραγματεύσεις της «κίτρινης» ομάδας με τη διοίκηση της Ξάνθης ξεκίνησαν (= le jaune est la couleur des équipes grecques AEK et ARIS)

[Les négociations de l'équipe « jaune » avec l'administration de Xanthi ont commencé]

[26] Τον κοίταζε με κίτρινα, πεθασμένα, αλλά διεισδυτικά μάτια (= le jaune symbolise la pâleur)

[Il le regardait avec des yeux jaunes, morts, mais pénétrants]

[27] Προειδοποίησε την Ελλάδα πως βρίσκεται στην «κίτρινη φάση» (= le jaune symbolise le degré d'organisation dans lequel se trouve le pays organisateur des Jeux Olympiques).

[Il a averti la Grèce qu'elle se trouve à la phase « jaune »]

Parmi les occurrences du *corpus* figurent des composés polylexicaux ou des expressions figées¹⁰ avec un sens métaphorique : par ex. : κόκκινα φανάρια/*κόκκινο φανάρι [feux rouges] = bordel, είμαι το κόκκινο πανί/*είμαι τα κόκκινα πανιά [être le chiffon rouge] = provoquer les autres, πράσινα άλογα/*πράσινο άλογο [chevaux verts] = choses inexistantes.

[28] Το Βερολίνο δεν είχε γειτονιές με «κόκκινα φανάρια» όπως άλλες ευρωπαϊκές πρωτεύουσες

[Berlin n'avait pas de quartiers avec des « feux rouges » comme d'autres capitales européennes]¹¹

[29] Τα σχέδια της εταιρείας αποτελούν κόκκινο πανί για τους 30 000 εργαζόμενους

[Les plans de la société c'est le chiffon rouge pour les 30 000 travailleurs]¹²

[30] Ποιος Ισοκράτης και πράσινα άλογα;

[Quel Isocrates et des chevaux verts ?]

Les adjectifs πράσινος et κόκκινος sont largement utilisés dans des expressions :

– περιμένω/παίρνω/λαμβάνω/δίνω/ανάβω (το) πράσινο φως [attendre/prendre/donner/allumer (le) feu vert] :

[31] Ο τελευταίος περιμένει το «πράσινο φως» από τον πρωθυπουργό

[Le dernier attend le « feu vert » de la part du premier ministre]

[32] Η εταιρεία έχει πάρει το πράσινο φως από το Χρηματιστήριο

[La compagnie a obtenu le feu vert par la bourse]

[33] Έλαβε χθες το «πράσινο φως» από το Εθνικό Συμβούλιο Ραδιοτηλεόρασης

¹⁰ On constate l'interdiction de changement en nombre de ces formes.

¹¹ « Berlin n'avait pas de quartiers avec des bordels comme d'autres capitales européennes ».

¹² « Les plans de la société provoquent les 30 000 travailleurs ».

[Il a obtenu hier le feu vert par le Conseil national de radiotélévision]

[34] Έδωσε σε όσους καλλιτέχνες ήθελαν το *πράσινο* φως να βγάλουν τραγούδια τους σε Mp3

[Il a donné le feu vert aux artistes qui le désiraient de sortir leurs chansons en mp3].

[35] Είναι έτοιμος να ανάψει το «*πράσινο* φως» στη διοίκηση του Ολυμπιακού

[Il est prêt à allumer le « feu vert » à l'administration de Olympiakos] – ανάβω (το) κόκκινο φως [allumer (le) feu rouge] :

[36] Η παραίτηση του Νούνιεθ άναψε *κόκκινο* φως για τη μετακίνηση του Ντένισον

[La démission de Nuñez a allumé le feu rouge pour la mutation de Denison]

Il y a aussi des exemples rares comme les suivants :

[37] Ήταν ένας νεαρός σε άθλια κατάσταση. *Κατακίτρινος*, με έντονα τα σημάδια τραυματισμού (1 fois)

[Il était un jeune homme dans un triste état. Tout jaune, avec les marques éclatantes des blessures]

[38] Στο τσαντίρι της, στριγγλιάρα γύφτισσα, η *πρασινοκίτρινη* αμαζόνα του θανάτου, η Επιδημία (1 fois)

[Sous sa tente, gitane hargneuse, l'amazone verte et jaune de la mort, l'Épidémie]

[39] Έγιναν *φαιοκόκκινος* σοσιαλφασισμός (1 fois)

[Ils sont devenus du socialisme et fascisme gris et rouge]

[40] Με τέσσερις γλάρους και έναν *μπρούντζινο* ήλιο, «φωτίζει» τον δρόμο του επισκέπτη (1 fois)

[Avec quatre mouettes et un soleil de bronze, il « illumine » la voie du visiteur]

[41] Σπας για χρόνια το κεφάλι σου στη *μαρμάρινη* αυτοπεποίθησή σου (1 fois)

[Tu te cognes la tête pendant des années contre ta confiance de marbre]

[42] Και μια κραυγή στο στόμα μου. Μια *κατακόκκινη* κραυγή. (3 fois)

[Et un cri dans ma bouche. Un cri tout rouge].

Il faut souligner que les circonstances du discours sont déterminantes pour l'interprétation adéquate. Nous retrouvons certaines ambiguïtés où les mots entourant l'adjectif ne sont pas suffisants pour lever l'ambiguïté et, par conséquent, il faut étudier le contexte plus large.

Dans les phrases suivantes, *πράσινος* « vert » pourrait prendre soit le sens de l'équipe Panathinaïkos ou d'une autre équipe portant la couleur verte soit le sens

du parti socialiste grec. C'est le contexte plus large qui permet la désambiguïsation :

[43] Τα τελευταία χρόνια στο πράσινο στρατόπεδο το πρόβλημα είναι ενδοοικογενειακό

[Les dernières années dans le camp vert le problème reste dans la même famille]

[44] Πανηγυρισμοί : «πράσινο» ξεφάντωμα σε όλη τη χώρα

[Des festivités : Du défolement « vert » dans tout le pays]

[45] Στην ιστορική και αγαπημένη γωνιά της «πράσινης» οικογένειας

[Au coin historique et aimé de la famille « verte »].

5. LA DIDACTIQUE DES ADJECTIFS EN – INOS

Pour la découverte de l'interprétation intensionnelle des noms de matériau, nous proposons les trois stratégies de compréhension ci-dessous :

– la prise en compte du contexte. En utilisant des indices linguistiques internes et externes ¹³, l'apprenant peut inférer le sens des mots nouveaux pour lui. Or, et c'est là l'essentiel de notre proposition, qui, selon nous, ne pose pas de problèmes de faisabilité (Galisson 1980 : 137), nous proposons d'exploiter la polysémie systématique ¹⁴ des noms de matériau ainsi que des adjectifs dénominaux correspondants. De même, l'enseignant aurait intérêt à chercher si cette polysémie des noms de matériau existe dans la langue maternelle des apprenants (Galisson 1980 : 49, 64-68) ;

– le « devinement » (Tréville et Duquette 1996 : 69), auquel procède l'apprenant après avoir pris en compte des indices linguistiques internes ;

– la segmentation en morphèmes, une stratégie d'apprentissage directe de type cognitif (analysing expressions) (Oxford 1990 : 83). L'apprentissage des règles de formation de mots ainsi que de quelques dizaines de mots-clés et des préfixes et suffixes les plus fréquents en GM contribue à l'acquisition d'un très grand nombre de mots dérivés et composés. C'est pour cette raison que nous proposons l'enseignement systématique des préfixes et suffixes les plus importants à travers un grand nombre d'exercices variés (v. aussi Anastassiadis-Syméonidis 1996). L'analyse morphologique est aussi proposée comme une stratégie d'approche du sens d'un mot inconnu (Nation 1990, Carter 1998) ¹⁵. De même Schmitt (1997 : 207) distingue, d'une part, les activités qui sont adéquates à l'approche initiale et, de l'autre, les activités de consolidation. Les premières constituent des stratégies de

¹³ Selon Sternberg (1987) (chez Tréville et Duquette 1996 : 73) les suffixes p. ex. sont des indices linguistiques internes, et le fait que *acier* réfère à un métal constitue un indice linguistique externe.

¹⁴ Bianchi (2001 : 102) parle de polysémie systématique quand le processus de génération du sens étendu est un processus systématique et productif, en tant qu'il génère un sens étendu pour toute autre expression possédant un sens primaire du même type.

¹⁵ Évidemment dans les cas où le sens est compositionnel.

localisation et de découverte du sens lexical par inférence à partir de la structure morphologique du mot, du contexte et de la description livrée dans les dictionnaires.

En suivant Tréville et Duquette (1996 : 115), nous proposons au niveau débutant, parmi les matériaux sociaux (Galisson 1980 : 124), la sélection d'un texte contenant un grand nombre de noms de matériau à interprétation extensionnelle et intensionnelle et des activités où les unités lexicales sont d'abord présentées dans leur contexte. Par la suite, l'enseignant peut proposer des exercices en dehors du contexte, dans le but d'enrichir le vocabulaire des apprenants tout en utilisant l'analyse morphologique comme stratégie de compréhension. C'est alors qu'il peut utiliser des techniques d'analyse et de traitement de morphèmes, dans le but d'inciter les apprenants à utiliser de façon généralisée les règles de construction de mots, par ex. *ατσάλι* « acier » → *ατσάλ-ινος/ατσαλ-ένιος* « d'acier ». Enfin l'enseignant termine son cours en replaçant les mots dans des contextes adéquats, de façon que l'apprenant en déduise la signification de l'énoncé, en mettant en marche des stratégies d'inférence.

Dans le cas où le nom recteur de l'adjectif en *-inos* appartient à une classe prédictible à taux élevé, tout en respectant les besoins langagiers des apprenants, on devrait donner la priorité à ces noms recteurs et procéder par regroupements. On pourrait enseigner p. ex. *βελούδιος* « de velours » de la manière suivante : *βελούδιος* + [SON] ¹⁶

φωνή	voix
ήχος	son
ψίθυρος	murmure
<i>βελούδιος</i> + [RENVERSEMENT]	
επανάσταση	révolution
ανατροπή	renversement
διαζύγιο	divorce

¹⁶ Les unités en majuscules représentent des concepts.

βελούδιος	+ [SURFACE]
επιφάνεια	surface
επιδερμίδα	peau
υφή του δέρματος	texture de la peau
βελούδιος	+ [LIQUIDE À CONSOMMER]
ντιπ	dip
σούπα	soupe

On évite l'enseignement des noms recteurs qui sont des métaphores vives, dues à l'originalité du locuteur, au moins au niveau débutant et moyen.

6. CONCLUSION

Le contexte linguistique et extralinguistique permet à l'auditeur/lecteur de formuler des hypothèses adéquates pour aboutir aux interprétations désirées. Sauf les cas d'ambiguïté, l'auditeur/lecteur peut arriver à interpréter le texte par inférence à partir des informations livrées par la situation extralinguistique.

Nous espérons que notre étude a démontré qu'une analyse à l'aide des *corpus* en général et des concordances en particulier peut mieux étayer une recherche théorique et arriver à des conclusions originales et utiles. Un dernier point intéressant réside dans le fait que l'application de cette méthodologie dans le domaine de la recherche sur la métaphore pourrait éventuellement inspirer des recherches dans le domaine de la polysémie systématique.

BIBLIOGRAPHIE

- Anastassiadis-Syméonidis (A.), 1987 : « Noms de couleur en grec moderne. Contribution lexicologique » dans *Studies in Greek Linguistics. Proceedings of the 8th Annual Meeting of the Department of Linguistics*, Faculty of Philosophy, Aristotle University of Thessaloniki, pp. 391-413. [En grec]
- Anastassiadis-Syméonidis (A.), 1996 : « La didactique des mots construits » dans *Le grec moderne comme langue étrangère*, Fondation Goulandris-Horn, pp. 129-149.
- Anastassiadis-Syméonidis (A.), 1997 : *Introduction à la morphologie dérivationnelle du grec moderne – I. Les adjectifs dénominatifs*, Thessaloniki.
- Anastassiadis-Syméonidis (A.), 2004 : « Que doit la compréhension au contexte et au registre de langue ? », dans Castagne (E.), édité, *Intercompréhension et Inférences*, Presses universitaires de Reims, pp. 21-39.
- Bianchi (C.), 2001 : « La flexibilité sémantique : une approche critique » dans *Langue française*, 129, pp. 91-110.
- Cadiot (P.) et Visetti (Y.-M.), 2001 : *Pour une théorie des formes sémantiques*, Paris, PUF.

- Efthymiou (A.), 1999 : *Le suffixe-ia en grec moderne. La manifestation d'un degré maximal d'anthropocentricité*, thèse de doctorat, Université Lille III, PUL.
- Galisson (R.), 1980 : *D'hier à aujourd'hui, la didactique des langues étrangères. Du structuralisme au fonctionnalisme*, Paris, Clé International.
- Kleiber (G.), (1994) : « Le problème de la déviance » dans *Langue française* 101, pp. 35-56.
- Kleiber (G.), (1999) : « Une métaphore qui ronronne n'est pas un chat heureux », dans Charbonnel et Kleiber, *La métaphore entre philosophie et rhétorique*, Paris, PUF.
- Marque-Pucheu (C.), 2001 : « Présentation » dans *Langue française*, 129, pp. 3-20.
- Noailly (M.), 1996 : « De l'image au concept : le nom en position d'épithète » dans Flaux (N.), Glatigny (M.), Samain (D.) édit., *Les noms abstraits. Histoire et théorie, Actes du colloque de Dunkerque 1992*, Ville-neuve d'Ascq, Presses Universitaires du Septentrion.
- Oxford (R.), 1990 : *Language learning strategies : what every teacher should know*, Boston, Heinle & Heinle.
- Prandi (M.), 1992 : *Grammaire philosophique des tropes*, Paris, Minuit.
- Salteri-Cacouros (V.) et Tamba (I.), 2001 : « L'interprétation métaphorique au prisme de la traduction : les constructions nominales du français, du grec et du japonais » dans *Langue française*, 129, pp. 54-63.
- Sinclair (J.), 1991 : *Corpus, concordance, collocation*, Oxford, Oxford University Press.
- Sperber (D.) et Wilson (D.), (1989) : *La Pertinence : communication et cognition*, Paris, Éditions de Minuit.
- Rastier (F.), 1987 : *Sémantique interprétative*. Paris, PUF.
- Tamine (J.), 1979 : « Métaphore et syntaxe », *Langages* 54, pp. 65-81.
- Tréville (M.-C.) et Duquette (L.), 1996 : *Enseigner le vocabulaire en classe de langue*. Paris, Hachette.

LES CONTEXTES : À LA SOURCE DU TERME ?

Cristina Alexandru et François Gaudin
laboratoire dynamiques sociolinguistiques
Université de Rouen
France

Résumé : *L'étude de quelques textes de vulgarisation scientifique nous permet d'étudier dans quelle mesure le cotexte détermine le sémantisme du terme et de quelle façon le contexte domaniale participe au processus de référénciation. Ces deux phénomènes relevant, pour les polysèmes, d'une sémantisation contextuelle.*

Nous montrons également comment la spécialisation des locuteurs dans la fixation de la référence impose une approche sociolinguistique du phénomène de la référénciation.

Mots-clés : *vulgarisation, référénciation, cotexte, division du travail linguistique.*

1. LE TERME : UN MOT EN CONTEXTE ?

Nous utiliserons dans cette contribution le terme *contexte* dans deux acceptions : celui de contexte textuel, ou *cotexte*, et celui de contexte pragmatique. Nous étudierons dans quelle mesure le premier détermine le sémantisme du terme et de quelle façon le second garantit le processus de référénciation, et ceci, en nous limitant au cas des textes qui nous intéressent, ceux relevant de la vulgarisation scientifique (désormais VS).

1.1. *Mot et terme* : une opposition ?

Les recherches récentes en terminologie ont montré que l'on ne peut plus opposer sans nuances mots et termes ; la critique de la monoréférencialité et l'univocité renvoient à un moment révolu. Nous nous demanderons donc aujourd'hui comment aborder du point de vue sémantique et référentiel cette unité théorique qu'est le terme et nous allons vous proposer modestement un possible modèle d'interprétation que nous exemplifierons sur un *corpus* d'articles provenant de la revue *Science & Vie*.

Aux commencements de la terminologie, ces deux notions – terme et mot – ont été constamment mises en opposition par Wüster et ses continuateurs ; d'une part, parce que la terminologie devait être délimitée de la lexicologie ; d'autre part, parce que le terme était envisagé comme un outil de communication, idéalement univoque et monoréférentiel, et dont le sens unique était établi par rapport à un domaine. Cette étape correspondait aux efforts épistémologiques d'un nouveau champ de recherche pour affirmer des fondements qui lui soient propres.

Avec les travaux de Teresa Cabré (1998), Valérie Delavigne (2001), Yves Gambier (1993 a et b), François Gaudin (2003) ou Monique Slodzian (1993), nous avons vu les apories de cette approche. Dans ce nouveau cadre, le contexte discursif a repris ses droits, longtemps mis de côté par la primauté du domaine. En admettant la polysémie du terme et la nécessité d'un contexte pour la désambiguïsation du sens, se pose désormais la question des nouvelles formes de détermination de la référence, et des nouveaux critères pour identifier ces mots dont le sens spécialisé est activé par et dans le contexte.

1.2. Terme et contexte : deux notions qui ne s'excluent plus

Pour peu que son usage gagne en fréquence, tout monosème est en danger de polysémie, et dès que nous parlons de polysémie, nous avons besoin de contextes pour actualiser le sens pertinent d'un mot ou d'un terme... polysémique. Hors du domaine, le terme a peut-être perdu son statut immuable, son sens unique en dehors du contexte, mais il a sûrement gagné en richesse sémantique, devenant aussi caméléon que le « simple » mot. Sa polysémie n'est plus à fuir car c'est un état de fait attribuable à l'entrecroisement des domaines, au mixage avec les discours quotidiens, à la dynamique de l'évolution de la science...

Notre hypothèse est qu'en dehors du domaine – domaine pensé comme régularité macrosémantique –, le terme a besoin pour être actualisé par le locuteur d'un *microcontexte linguistique* ; celui-ci est composé en général d'autres termes qui ensemble forment un domaine de référence, un champ conceptuel qui aide à situer du point de vue sémantique tous les composants. Nous proposons donc une analyse en termes de *sémantique relationnelle*, « des relations non pas dans le mot, mais entre les mots » (Lerat 1995 : 187) et cette précision importe dans la mesure où « l'une des caractéristiques les mieux assurées des termes, c'est sans doute qu'ils prennent toute leur valeur une fois situés dans un ensemble terminologique pertinent. » (Gaudin 2003 : 151).

Or, justement cet ensemble terminologique pertinent se forme dans un contexte qui fournit le cadre de référence. Dans nos exemples, nous avons sélectionné des cas où le cotexte, c'est-à-dire les mots situés dans la proximité immédiate d'un terme, forme un champ conceptuel qui permet d'établir correctement le sens des mots et de construire une impression référentielle. Le terme se retrouve entouré d'autres unités terminologiques, et cet ensemble dessine un champ sémantique qui nous indique le domaine d'activité vers lequel nous diriger pour actualiser correctement le référent extralinguistique d'un terme en même temps que son sens par rapport à son entourage linguistique.

Dans notre étude, nous nous sommes intéressés à des termes dont l'usage fréquent dans la langue a conduit à l'*acculturation* des terminologies, phénomène qui s'accompagne naturellement d'une *banalisation* des termes (cf. Galisson 1978). C'est le cas des mots provenant pour la plupart de l'univers technologique, tels *caméscope*, *imprimante*, *micro-ondes*, *photocopieuse* ; et de la médecine, tels *appendicite*, *aspirine*, *cancer*, etc. De même, nous observons des mots polysémiques appartenant à la langue générale qui, dans un domaine scientifique, développent

des sens particuliers. C'est le cas des mots comme *collision*, *croûte*, *pli*, *s'étirer* que nous allons analyser en ce qui suit.

Dans le discours de VS, le terme gagne en fréquence, en souplesse sémantique (il s'enrichit grâce à la multiplicité des discours où il apparaît), mais il perd en précision, en substance théorique, — on peut dire de la rigueur scientifique qui le caractérise dans le discours des scientifiques. Les conditions pragmatiques changent, le sens des mots aussi ; les référents extralinguistiques semblent les mêmes, mais les façons d'y renvoyer sont-elles identiques ? Nous y reviendrons.

La vulgarisation se situe dans cette aire de la communication, par son statut même de discours de popularisation de la science. Définir tous les termes employés le long d'un article serait non seulement difficile, mais absurde. Il y a dans la VS une grande partie d'implicite, mais aussi beaucoup de termes dont les significations s'accomplissent grâce au contexte. La définition est rare comme pratique car c'est une forme métalinguistique contraignante. Le contexte définitoire présente l'avantage de s'inscrire dans les visées pragmatiques du discours et d'être adapté aux lecteurs.

La VS est une pratique d'écriture complexe qui favorise la circulation des termes et l'acculturation terminologique. Quel sens pour les termes dans ces discours ? Sont-ils les mêmes que dans les laboratoires où ils se forgent ? Et, une fois repris dans les discours quotidiens, quelles fonctions remplissent-ils : communiquer un savoir ? Étiqueter des objets techniques sophistiqués qui nous entourent ? Ou connoter la scientificité ? Des mots comme *micro-ondes*, *pyrolyse*, *induction* sont-ils plutôt des mots ou des termes ? Voici autant de questions qui sont au cœur du débat sur le statut sémantique des termes en contextes variés...

1.3. Quelques exemples

Étant donné le lectorat qu'elles visent, les revues de VS se trouvent dans la situation d'expliquer, de populariser des théories scientifiques, les dernières découvertes, de rendre la science plus accessible. Cette accessibilité est travaillée à travers une rhétorique qui recourt à des images, des explications, des paraboles, des parallélismes avec la vie quotidienne, etc. Le répertoire des procédés utilisés est très vaste, même si, bien sûr, tous les termes ne font pas l'objet d'une définition ou d'une reformulation.

Nous proposons donc l'analyse de quelques mots dont le sens terminologique ne s'actualise que dans certains contextes ; il s'agit de mots polysémiques, qui voient leur sens spécialisé actualisé grâce à leur contexte. Nous prendrons tout d'abord les cas de *collision* et de *croûte*, pris dans le même article.

Pour le *Dictionnaire historique de la langue française*, *collision* « désigne un choc entre deux corps en mouvement ». Il a reçu de nombreuses valeurs figurées, spécialement en musique et en linguistique ainsi qu'en physique nucléaire, pour « choc entre particules ». Pour sa part, le *Trésor de la langue française informatisé* énumère des emplois spécialisés en linguistique, en musique et dans le secteur du nucléaire, emplois qui tous évoquent l'idée de choc, de contact direct, mais pas

d'emploi spécialisé pour la géologie. En revanche, le *Grand Dictionnaire terminologique*¹ enregistre, à côté des définitions pour l'aéronautique, les assurances, le chemin de fer, industrie minière, informatique, marine, physique, sport et télécommunications, un emploi spécialisé pour la géologie, défini ainsi : « Choc entre deux masses continentales résultant de la fermeture d'un domaine océanique intermédiaire, et s'accompagnant de déformations très importantes (orogénèse, chaîne de collision). »

L'article qui nous sert d'exemple, « Alpes : l'étrange crise de croissance » (*Science & Vie* juin 2005) autorise l'actualisation d'un tel sens ; pourtant nous ne retrouvons pas de définition comme celle que nous venons de citer dans l'article. La compréhension n'en est pas pour autant gênée, justement parce que l'idée générale de choc est gardée ; mais si nous voulons connaître la signification de ce mot par rapport à la géologie, nous faisons appel à un expert, le dictionnaire, qui nous fournit les détails pour avoir une référence complète. Ce qui est un exemple de la *division du travail linguistique* dont nous reparlerons.

Le thème de l'article oriente d'emblée l'interprétation des polysèmes : des mots comme *collision*, *croûte*, *chaîne*, *pli*, *faille*, *cassure*, *surveillance* (géodésique), *mesures* (géodésiques) *s'étirer*, *se raccourcir*, que l'on comprend facilement en dehors du discours, forment un ensemble lexical qui, rapporté au domaine de la géologie, fait apparaître des sens spécialisés. Le sens général de chacun de ces mots pris comme exemple est conservé, mais le domaine sémantique – dans notre cas, la géologie – fait apparaître des nuances et des particularités pour l'emploi spécialisé.

Nous retrouvons dans ce texte un cas intéressant de synonymie discursive ; pour deux termes composés nous sommes renvoyés à une même explication, appartenant à la rubrique « Jargon ». Ainsi les dénominations *mesures géodésiques* et *surveillance géodésique* deviennent des termes synonymes de discours comme le confirme la définition donnée :

Mesures géodésiques : *Elles permettent d'évaluer les déformations de la croûte terrestre. Quand elles sont réalisées grâce au système de positionnement par satellite (GPS), elles peuvent mettre en évidence des déplacements de seulement quelques millimètres.*

Nous sommes donc en présence d'une synonymie discursive, pertinente dans la mesure où le discours traduit une expérience collective, celle des géologues.

Cette expérience se concrétise dans le *contexte élargi* représenté par l'article ; le sujet dont il traite fonctionne comme un *espace mental* (Fauconnier 1983) qui actualise l'emploi terminologique des mots comme ceux que nous avons pris pour exemple ; c'est pourquoi, à la lecture, nous comprenons ces mots à travers le domaine de la géologie, excluant, par exemple un emploi figuré pour le mot *faille*, ou le sens pictural de *croûte*. Ces lexies, polysèmes pour le descripteur linguiste, prennent un sens spécialisé une fois actualisées dans un discours scientifique qui

¹ www.granddictionnaire.com

conduit à la formation d'un champ lexico-sémantique. Ce phénomène relève en partie de l'isotopie, phénomène qui se manifeste souvent, dans la vulgarisation, par des séries énumératives de mots qui tous actualisent, grâce au contexte, un sème commun.

Un autre article du même numéro de *Science & Vie*, traitant des maladies cardiovasculaires, en offre des exemples :

« Pourtant, l'excès de cholestérol n'est qu'un coupable parmi... plus de 300 autres identifiés à ce jour ! Parmi eux : le fait d'être un homme (les jeunes femmes sont moins touchées, protégées par les œstrogènes), l'âge (plus on vieillit plus nos artères s'abîment), et l'hérédité, trois facteurs sur lesquels on ne peut agir. S'ajoutent des facteurs modifiables ; hypertension, diabète sucré, tabac, obésité abdominale, alimentation, sédentarité, alcool et stress. »²

Les trois premiers mots soulignés actualisent ici des sens particuliers par rapport à la santé ; si l'âge et l'hérédité peuvent être définis en tant que *facteurs* qui favorisent la sensibilité à l'excès de cholestérol, *homme* que l'on oppose à *femme*, acquiert aussi cette qualité de facteur qui favorise certains comportements de l'organisme. À partir du cotexte qui favorise des oppositions discursives, nous pouvons définir les mots de notre série en usage médical :

- *homme* : peut être plus facilement atteint par l'accumulation en excès du cholestérol ;
- *âge* : un des facteurs qui causent la détérioration des artères ;
- *l'hérédité* : favorise la sensibilité aux excès de cholestérol.

Le contexte immédiat a rendu possible l'élaboration de ces définitions ; en dehors de ce contexte, *homme* est plutôt défini comme l'opposé de *femme*, l'âge comme durée de la vie humaine et l'hérédité comme transmission fondée sur les liens du sang.

Dans la série suivante, nous avons mis en évidence le comportement sémantique des mots *tabac*, *alimentation*, *alcool*. En dehors du contexte où nous les retrouvons, le sens de ces mots est *tabac* : « Produit manufacturé, vendu sous diverses formes, fabriqué à partir des feuilles de tabac séchées et préparées pour fumer, priser ou chiquer. » (TLFI) ; *alimentation* : « Ce qui alimente, produits alimentaires » (TLFI) ; *alcool* : « Boisson contenant un excitant issu de la fermentation et pouvant être isolé par distillation. » (TLFI) Le *Grand Dictionnaire terminologique* (TLFI).

Malgré une consultation des articles du TLFI et du *Grand Dictionnaire terminologique*, nous n'avons pas retrouvé de définition liée au domaine de la santé. Comme dans les exemples précédents, les mots analysés acquièrent leurs sens dans le discours et grâce au contexte immédiat qui fait que toute la série de mots cités actualisent un sens lié à la santé, un sème afférent, encore plus précis, lié au problème du cholestérol. Ces définitions accomplies en discours et d'ordre plutôt encyclopédique, nous ne les retrouvons pas dans les dictionnaires. Ce sont des valeurs

² Nous soulignons.

acquises en discours et comme ce discours tient des domaines de la médecine et santé, nous pouvons les enregistrer comme des emplois discursifs spécialisés.

Analyser les termes en contexte fournit des informations sémantiques auxquelles on ne pourrait pas accéder en dehors de ces manifestations discursives. Ceci est d'autant plus important quand les mots sont des polysèmes activant leur caractère de terme en discours et quand les discours appartiennent à la VS.

Dans ce type de textes, les mots-termes possèdent une fonction communicative qui favorise encore plus le développement de la polysémie ; le cotexte et le domaine sémantique évitent la surcharge du texte avec des définitions, dont la présence massive serait d'ailleurs ressentie comme artificielle ; les isotopies créent des régularités sémantiques, non enregistrées dans les dictionnaires, qui peuvent conduire à la construction de nouveaux concepts.

Enfin, si le cotexte et le contexte domaniale, permettent de construire une signification acceptable en rapport avec les besoins pragmatiques du lecteur, ce dernier reste dépendant pour la validité de la référénciation des vocables, d'un expert qui puisse garantir le lien entre un nom et une classe de référents.

2. DE LA RÉFÉRENCE COMME PHÉNOMÈNE SOCIOLINGUISTIQUE

Ce qui précède concerne le cotexte, l'entour textuel qui permet de dire qu'une forme lexicale prend plutôt un statut de terme que de mot. Dans cette seconde section, nous voudrions examiner le rôle du contexte éditorial dans la spécificité du terme du point de vue référentiel.

2.1 Référenciation et déférence

Le plus souvent, la référence est appréhendée en terminologie comme inhérente au signe. C'est une propriété sémiotique et non extra-sémiotique comme chez Saussure. Cet état de choses se comprend dans la mesure où, ce qui prime, c'est l'efficace du signe, sa force de renvoi vers le monde et non l'infinie virtualité de sa puissance signifiante.

Ordinairement, l'usage des noms se passe dans l'accord, grâce à la convention, et, si besoin est, la négociation ; la référenciation est alors mutuellement consentie ou construite. Mais que se passe-t-il si un doute ou un désaccord survient sur le lien entre un signe et l'expérience du monde : « tu es sûr que ça s'appelle un *cross-over*³, une carambole⁴, une épaufrure⁵ ? »

Dans ces cas, il nous faut recourir à une personne ressource qui soit garante, dans notre communauté linguistique, de la validité du lien entre nom et classe référentielle. Les auteurs de vulgarisation font partie de ces personnes-ressources. Les

³ Un *cross-over* est un épisode exceptionnel dans lequel les personnages de deux séries différentes se retrouvent au sein d'une même intrigue, ce que ne rend ni *liaison* ni *croisement*.

⁴ Baie exotique comestible.

⁵ Éclat emporté sur une arête d'une brique, d'une pierre, etc.

lecteurs sont de l'autre côté ; ils n'ont pas légitimité pour décider si ce que l'on voit à la télévision est un *cross-over*, si le détail d'un mur est une épaufrure, si le métal de la bague est bien de l'or blanc, si les chaînes d'acides aminés sont bien des protéines.

La vulgarisation est donc un bon exemple de ce que le philosophe Hilary Putnam appelle la *division du travail linguistique*. Cette idée a été avancée pour souligner que le contenu sémantique ne dépendait pas uniquement des individus, mais, assez largement, des ressources cognitives présentes dans un environnement tant physique que social. Pour savoir vraiment ce qu'est une carambole, un *cross-over*, une épaufrure, une protéine, une psychose, je m'adresse à des personnes ressources, le plus souvent par le biais d'écrits, c'est ce que les philosophes du langage appellent la *déférence*.

Or ce que nous voudrions faire ici, c'est insister sur le fait qu'il s'agit d'un phénomène sociolinguistique, car comme le souligne Putnam lui-même, ce phénomène repose en partie sur la division du travail, donc sur un fait social. La division du travail impose, pour les termes, le phénomène de la déférence. Or si la référence fait l'objet d'une compétence socialement répartie, cela met en cause l'homogénéité du fonctionnement des signes. Certes, ce phénomène ne concerne pas tous les noms ; même s'il existe des menuisiers et des ébénistes, utiliser des noms tels que *table* ne pose pas de problème. Mais à quoi référons-nous quand nous parlons de protéines, d'une eau calcaire, de *Pytiriasis versicolor* ? Si ces signes sont pourvus d'une référence socialement distribuée, cette référence peut-elle être considérée comme une propriété de ces termes ? Et ne doit-on envisager une approche sociolinguistique de la référence si celle-ci cesse d'être une propriété sémiotique du terme ?

L'approche de Putnam permet de replacer dans le circuit des interactions verbales, donc de la communication réelle, des notions souvent abordées de manière idéaliste. Nous avons tous fait l'expérience que, pour certains noms, nous sommes en relation avec la référence de certains des termes de notre lexique, non individuellement, mais via la distribution des connaissances dans notre communauté linguistique. Les autres savent pour moi ce que sont exactement une chaîne de distribution, une chanterelle, un cromorne, ou un écoulement diphasique...

Ordinairement, notre contenu intentionnel, notre intention de référer, suffit ; mais si nous avons un doute sur le nom d'un champignon, nous allons voir le pharmacien par un mouvement de déférence. Les deux relèvent deux ordres différents : l'intention de référer relève de la représentation et le savoir du pharmacien relève du contenu épistémique. Il est évident que les problèmes d'expertise, de spécialisation relèvent de cette division du travail linguistique. Cette division est constituée de deux pôles, mais il existe un *continuum* – heureusement ! Entre les deux, le langage joue un rôle unificateur. Et la VS s'inscrit au sein de ce continuum qu'elle contribue à construire.

Cette opposition entre significations ordinaires et significations savantes peut être reliée à l'opposition entre représentations et concepts, entre ce qui relève de la psy-

chologie individuelle et de la logique. Mais entre les deux, existe un dialogue, un continuum, et chacun d'entre nous manipule à la fois des représentations et des concepts. Si l'on est très préoccupé par l'exactitude, on dira avec Bachelard que cela peut constituer un obstacle pour la connaissance, car c'est une des sources possibles d'erreur, mais cela constitue un heureux avantage pour la communication humaine.

2.2. Une référencement socialement garantie

Revenons à la vulgarisation pour insister sur le lien existant entre contexte et référence.

La vulgarisation constitue le type d'interaction dans lequel existe une différence entre deux membres de la communication. Il y a donc division du travail linguistique : le lecteur s'en remet à l'auteur pour ce qui est de la vérité de ses affirmations et il le suit pour construire de représentations qui lui permettent une référencement. Or il ne s'agit pas pour le lecteur d'une simple confiance ordinaire ; il existe une instance qui relève du contexte : la vulgarisation scientifique constitue un contexte éditorial, et donc social, qui garantit la scientificité de l'auteur et donc de ses propos. Cette garantie passe par le truchement de conseils scientifiques, de comités éditoriaux, qui cautionnent l'appartenance des auteurs au côté autorisé de la division du travail linguistique. De ce point de vue, les sciences sont intéressantes à observer car il y a une frontière assez aisée à tracer entre ce qui relève socialement de l'institution scientifique. Tous les chercheurs, tous les enseignants-chercheurs présents dans une salle sont réputés scientifiques ; tous les journalistes, tous les étudiants, tous les pompiers ne le sont pas. Et un chercheur journaliste ne le sera qu'une partie de la journée !

Or cette spécialisation sociale fait partie du régime qui caractérise le fonctionnement référentiel des termes. On pensait qu'il s'agissait d'un régime possédant la noblesse philosophique de l'ontologie, on se retrouve avec un régime relevant du social ; ou si l'on veut des fonctionnements sociolinguistiques.

Si l'on suit notre raisonnement, la spécificité du terme, si elle existe, n'est ni dans le signe, ni dans le domaine, mais dans le fonctionnement socialement réparti de la référencement, par le biais de ce que l'on appelle la *déférence*. Le contexte social et éditorial indique que l'on se tient du bon côté de la division du travail linguistique, du côté des autorités référentielles. Et la vulgarisation permet au lecteur d'entrer dans les coulisses du monde de ceux qui savent si la bague est en or, si le fruit est bien une carambole, si le creux de la pierre est une épaufrure et si l'apparition de Columbo dans un épisode de Derrick est un *cross-over*. Et les coulisses sont un lieu privilégié entre les lumières de la scène et l'ombre de la foule.

3. CONCLUSION

Si la sémantisation du terme est garantie par le cotexte, donc par son régime textuel, la référencement relève de son régime sociolinguistique.

Les références savantes sont donc entre les mains des experts, comme le dit la terminologie depuis toujours. Et au fond, ce que cherchait en premier lieu la terminologie originelle, c'était à améliorer les échanges de ce côté-là de la communication en permettant un rapport stable entre dénominations et significations (voir Candel 2004). Mais l'évolution des sociétés modernes a rendu nécessaire une interrogation sur la diffusion des savoirs scientifiques qui relève de l'autodidaxie. Ce qui suppose une souplesse dans une société de savoirs répartis et circulants.

Or on sait depuis Jacotot que l'autodidaxie a partie liée à l'émancipation et qu'elle s'oppose à l'instruction qui reconduit une séparation des rôles dans la circulation des savoirs (sur Jacotot, voir Rancière, 1987). La langue est là pour tisser des liens entre les représentations des profanes et les concepts des experts, de ceux auxquels on défère. Ce dialogue est une fonction essentielle de la VS. De ce point de vue, observer les processus qu'elle met en jeu, c'est chercher à comprendre comment le lecteur peut s'aider lui-même à briser la vitre derrière laquelle le tient son ignorance.

BIBLIOGRAPHIE

- Alexandru (C.), 2006, *Taratata*, thèse de doctorat, Universités de Bucarest et de Rouen, Cunita et Gaudin (dir.).
- Cabré (M. T.), 1998 : *La Terminologie. Théorie, Méthode et Applications*, Ottawa-Paris, Presses de l'Université d'Ottawa et Armand Colin, 322 pp.
- Candel (D.), 2004 : « Wüster par lui-même », *Cahier du Ciel* 2004, pp. 15-31.
- Delavigne (V.), 2001 : *Les Mots du nucléaire. Contribution socioterminologique à une analyse de discours de vulgarisation*, thèse de doctorat N.R., dir. F. Gaudin, Université de Rouen, 2 vol. 674 pp. + annexes.
- Fauconnier (G.), 1984 : *Espaces Mentaux : aspects de la construction du sens dans les langues naturelles*, Paris, Éditions de Minuit.
- Galisson (R.), 1978 : *Recherches de lexicologie descriptive : la banalisation lexicale*, Paris, Nathan, collection « Université, Information, Formation », 432 pp.
- Gambier (Y.), 1994 : « Implications épistémologiques et méthodologiques de la socioterminologie », dans *Alfa*, Halifax, Universitas Dalhousiana, vol. VII/VIII, pp. 99-115.
- Gaudin (F.) 2003 : *Socioterminologie. Une approche sociolinguistique de la terminologie*, coll. « Champs linguistiques », Louvain-la-Neuve, Duculot, 286 pp.
- Laurendeau (P.), 1998 : « Théorie des opérations énonciatives et représentations : la référenciation », dans *Cahiers de praxématique*, n° 31, pp. 91-114.
- Maurice (L.), 2004 : « Aux frontières du concept et de la représentation », *Colloque jeunes chercheurs*, ENS LSH Lyon.
- Origgi (G.), 2005 : « Croyance, déférence et témoignage » à paraître dans Proust (J.) et Pacherie (E.), *Introduction à la philosophie des sciences cognitives*, Paris, Éditions de la Maison des sciences de l'homme [<http://gloriaoriggi.free.fr/origgicroyance.pdf>]
- Putnam (H.), 1990 : *Représentation et Réalité*, Paris, Gallimard, coll. « NRF Essais », 226 pp.
- Rancière (J.), 1987, *Le Maître ignorant. Cinq leçons sur l'émancipation intellectuelle*, Paris, Fayard, 233 pp.
- Slodzian (M.), 1993 : « La V.G.G.T. (Vienna General Theory of Terminology) et la Conception Scientifique du Monde », dans *Le langage et l'homme*, vol. XXVIII, n° 4, DeBoeck Université, pp. 223-232.

LE TROU LEXICAL ENTRE LANGUE ET DISCOURS : POUR UN STATUT DISCURSIF DU SILENCE

Mamadou Diakité
Université Cheikh Anta Diop
Dakar, Sénégal

Résumé : *Le silence n'a pas de statut propre. Quasiment ignoré en discours, il est, en langue, réduit aux règles du langage formel, recevant différentes appellations selon le niveau d'analyse. Nous regroupons ces termes sous le nom de « silence actualisé », en l'opposant au silence en discours ou « silence réalisé ». Lié aux enjeux de pouvoir, celui-ci peut, sous ses deux principales modalités – total et partiel – être soumis aux méthodes de l'analyse du discours.*

Mots-clés : *discours, pouvoir, silence actualisé, silence réalisé.*

1. INTRODUCTION

L'étude du langage semble exclure le silence de son champ. Le paradoxe est que les sciences du langage traitent beaucoup du silence. Ainsi parle-t-on, selon les disciplines, de trou lexical, d'implicite, d'ellipse... Tous ces faits sont inscriptibles dans une problématique de langue, car ils consistent en suppression d'unités formelles dont l'absence est codifiée par le système linguistique dans le cadre de la phrase ou d'une structure lexicale. À ce type de silence, nous donnerons le nom de « silence actualisé », en référence au fait que la langue actualise le langage. Malgré la diversité de ses manifestations, ce type de silence n'épuise pas la réalité du phénomène. En effet, en plus du silence de l'ignorance et du vide sémantique que tente de remplir la fonction phatique du langage, il existe ce que Pulcinelli (1991 : 209) appelle le « silence fondateur », dont la fonction est de créer « l'espace différentiel » indispensable à l'existence du sens.

L'objet de notre propos est différent de tous ces types de silence. Il s'agit d'un phénomène discursif très peu étudié, et qui a partie liée avec les enjeux de pouvoir en général. Nous l'appellerons « silence réalisé », le discours étant la matérialisation de la langue. Notre objectif est d'en examiner le statut et le fonctionnement, en nous fondant sur l'idée que le discours est un processus sémantique constitué d'unités formelles et de silences et que partant, une procédure d'analyse du discours doit pouvoir intégrer le silence. Après des préliminaires portant sur des considérations théoriques, nous présenterons les occurrences de silence réalisé avant de les analyser à partir des contextes social et discursif. Notre *corpus* est constitué d'articles parus entre 1988 et 1997 dans deux quotidiens dakarois, SL (*Le Soleil*) et WF (*Wal fadjri*) et portant sur le DOP (discours sur l'optimum de popula-

tion) au Sénégal, discours dont la planification familiale est la partie visible et conflictuelle.

2. PRÉLIMINAIRES

Postuler un statut discursif du silence suppose qu'on peut lui appliquer les méthodes de l'analyse du discours dont se réclame la présente communication. Il serait donc utile de définir les concepts dont nous aurons besoin.

Par discours, nous entendons un processus sémantique, mot ou plus, rapporté à ces conditions de production. On l'appelle souvent formation discursive à la suite de Foucauld (1969), pour tenir compte de son caractère dynamique et donc de l'instabilité de ses limites qui se déplacent au gré des enjeux sociaux, politiques, idéologiques, etc. Ces frontières, appelées interdiscours, sont un espace d'interaction entre formations discursives où s'opère, pour chaque formation, le processus d'assimilation de composants extérieurs d'une part, le réaménagement de ses composants initiaux, d'autre part.

La notion d'interdiscours est impliquée par celle d'univers discursif¹, « l'ensemble des formations discursives de tous types qui [...] interagissent dans une conjoncture » (Maingueneau 1987 : 85). Cette entité se compose de champs² discursifs, espaces socioinstitutionnels relativement autonomes (politique, scientifique, religieux...) dans lesquels interagissent des formations discursives à travers des processus sémantiques, dont le silence.

Nous avons déjà dit que le silence actualisé consistait en effacement d'unités formelles. Le silence réalisé aussi peut porter sur des unités lexicales, mais à un niveau macrostructurel, puisqu'il résulte d'enjeux sociaux liant des formations discursives. À ce titre, le silence réalisé peut être conçu comme un trou lexical qui relève non pas de la langue, mais du discours. En plus de cette différence, et contrairement au silence actualisé, le silence réalisé peut effacer un objet socio-discursif global ou thème : c'est dire finalement qu'il a une portée. Nous parlerons de silence partiel ou total selon que l'effacement est circonscrit à l'unité lexicale ou porte sur un thème.

Latent en langue, le silence est, en discours, l'expression d'un rapport de forces résultant, selon le mot de Pulcinelli (1991 : 208) d'un « processus de mise-en-silence [ou] “silenciation” ». Il est différent du « non-dit qui – tel l'implicite – vient s'ajouter au dit mais il est un non-dit nécessairement effacé, écarté du dit. » Ainsi conçu, le silence réalisé est soumis aux mêmes contraintes que les discours qui le supportent. Dans tous les cas, il reste une forme de violence du pouvoir³, sa fonction étant de réduire l'autre à ce que Bourdieu (1998 : 24) appelle la « soumission paradoxale », effet de « la violence symbolique » qui s'exerce surtout par « les voies purement symboliques de la communication et de la connaissance — ou, plus

¹ L'univers discursif de Maingueneau correspond assez bien à l'interdiscours d'Amossy (2000 : 89-90).

² La notion de champ est l'un des concepts fondamentaux de la sociologie de Bourdieu.

³ Qu'on pense à l'expression « conspiration du silence ».

précisément, de la méconnaissance, de la reconnaissance ou, à la limite, du sentiment. » Ainsi, en Europe, pendant de longs siècles, la maîtrise du latin a constitué un instrument de violence car « Le pouvoir de dire » des « uns avait pour contrepartie le silence des autres, réduits de nécessité à la croyance et à l'obéissance, manifestant au plus la conscience de leur condition » (Waquet 1998 : 274).

Toutefois, si le savoir et le pouvoir-dire sont les principaux instruments de la violence symbolique, le rôle du sentiment ne doit pas être négligé pour autant. Celui-ci peut, dans certaines conditions et plus souvent qu'on ne le dit, neutraliser un savoir et un pouvoir-dire avérés. Son action est d'autant plus pernicieuse qu'elle se manifeste généralement sous la forme de la doxa. Bourdieu (1996 : 25) explique à ce propos comment, dans une émission de télévision, il a été contraint au silence, car placé devant l'alternative de se soumettre en se taisant ou de faire un scandale « contraire aux règles du débat “ démocratique ” ».

Le silence réalisé ne procède donc pas nécessairement de l'action du savoir sur l'ignorance. Les choses sont plus complexes : comme le discours dont il est un composant nécessaire, le silence, en dernière analyse, ne peut se comprendre indépendamment des interactions sociales qui l'ont produit. La construction de notre *corpus* tient compte de ce postulat.

3. LE *CORPUS*

3.1. Le contexte discursif

La procréation au Sénégal est fortement valorisée par les traditions et les religions (94 % de musulmans et 4 à 5 % de chrétiens), appuyées pendant soixante ans, par une loi française de 1920⁴ interdisant la contraception. L'abrogation de cette loi en 1980 a permis la pratique légale de la contraception mais de fortes pressions religieuses interdisaient d'en parler, si bien que le DOP n'est apparu qu'en 1988, date de l'adoption de la politique de population. Pourtant, de 1961 à 1985, un taux de croissance démographique supérieur au taux de croissance économique a entraîné l'appauvrissement de larges couches de la population. Toujours est-il que, sur le plan social, la naissance du DOP a été négociée sur la base de l'idée que l'islam détermine le comportement de ses adeptes sénégalais en matière de procréation.

Le DOP est investi par quatre champs discursifs : politique, religieux, celui de la tradition et celui de la presse. Le champ discursif politique comprend les formations discursives étatique et interétatique, en l'occurrence le FNUAP (Fonds des Nations unies pour la population⁵) et ses relais dans la société civile. Le champ discursif religieux est plus complexe. Sa principale composante est l'islam confrérique sunnite, numériquement dominé par les confréries *Mouride* et *Tidiane*. Cette formation est concurrencée depuis environ un quart de siècle par un mouvement

⁴ Le Sénégal est une ancienne colonie française.

⁵ Nous simplifions quelque peu en établissant une relation horizontale entre l'État et le FNUAP : en réalité, celui-ci est le commanditaire de celui-là. Nous les regroupons dans la même instance parce qu'ils ne s'opposent pas.

proche du chiisme iranien, le « nouveau courant islamique ⁶ ». Le champ discursif de la tradition a un statut problématique : les autres formations lui déniaient la moindre existence formelle en disqualifiant ses formations accusées de paganisme et de croyances obscures. Le champ de la presse comprend les deux quotidiens, SL et WF ⁷. Le premier est proche du pouvoir ; l'autre est un organe indépendant qui semble avoir beaucoup évolué, passant de 1984 à 1993, du statut d'organe islamique à celui de journal d'informations générales.

3.2. Présentation des occurrences

3.2.1. *Le silence total*

[1] Les 94 % de musulmans sénégalais sont répartis dans des confréries dont les plus importantes au plan numérique sont les mourides et les tidianes. Dans toutes les confréries — à l'exception notable des layènes, à partir de 1997 — le silence est général pour la période étudiée, alors que le nouveau courant islamique s'exprime largement sur le DOP depuis 1988.

[2] *Wal fadjri* observe un silence total de dix ans de janvier 1988 au 6 juin 1997. Son premier article sur la question paraît en 1997.

[3] Le discours de la tradition, dont le griot est l'interprète, connaît dix ans de silence avant de se manifester en 1997 (*Le Soleil*, 24/12/97).

3.2.2. *Le silence partiel*

Comme nous l'avons déjà dit, le silence partiel consiste en effacement d'une ou de plusieurs unités lexicales. Un tel phénomène est plus facilement repérable au niveau de la phrase que dans un *corpus* de près d'une centaine de milliers d'occurrences. Pour résoudre cette difficulté, nous avons procédé à une analyse lexicométrique fondée sur l'idée que le lexique n'est pas seulement une liste de mots, mais qu'il s'organise aussi en sous-systèmes. Le choix d'un sous-système ne doit pas cependant être arbitraire, il doit être explicable par le *corpus*. Dans cette perspective, nous avons constitué un *corpus* lexicométrique composé d'un échantillon d'articles ⁸ de chaque journal, en veillant à l'équilibre entre les journaux, tant du point de vue de la taille du *corpus* (nombre d'occurrences), que de la nature des articles et de la période étudiée ; soit au total 13 268 occurrences. Nous avons établi un index hiérarchique ⁹ pour le *corpus* global et retenu la fréquence 10 comme le minimum statistiquement significatif ¹⁰. La liste obtenue comporte, en plus de la

⁶ J'appelle ainsi le discours de l'intelligentsia arabophone formée pour l'essentiel dans les universités des pays arabes. *Wal fadjri* en a été l'un des initiateurs au début des années quatre-vingt.

⁷ La presse a un statut double : d'une part elle constitue un champ, d'autre part elle est le support pour tous les champs du *corpus*, y compris celui de la presse. Nous tenterons de tenir compte de cette ambiguïté dans l'analyse.

⁸ Voir les références à l'annexe 1.

⁹ Liste des occurrences classées par ordre décroissant de fréquence.

¹⁰ Ce choix est arbitraire, mais nous signalons que dans chaque journal, la fréquence 10 est la plus basse des cent formes les plus fréquentes.

fréquence totale (F) de chaque forme, sa sous-fréquence FSL ou FWF selon qu'il s'agit du sous-*corpus* SL ou WF.

Mais les fréquences ne sont pas intéressantes en soi car on sait que le thème, le style ou la taille du *corpus* les déterminent dans une large mesure ; c'est pourquoi nous avons porté notre attention sur la recherche de structures lexicales et avons alors constaté une disparité, remarquable pour un discours sur la procréation : la fréquence de « femme » est de 63 contre 14 pour « homme » (*vir*), soit 4,5 fois moins. Cette forte asymétrie a été considérée comme un demi-silence partiel, ce qui nous a conduit à faire l'hypothèse que le *corpus* est marqué par une problématique de genre, en l'occurrence le silence sur le genre homme. Ce fait est d'autant plus remarquable que, d'ordinaire, ce sont les femmes qui dénoncent « la conspiration du silence » dont elles seraient victimes. Pour vérifier cette hypothèse fondée sur l'examen des occurrences dont la fréquence est statistiquement significative, nous avons décidé d'examiner dans le reste du *corpus* lexicographique (formes de fréquence inférieure à 10) les cohyponymes de « femme » (mère, épouse, fille) et d'« homme » (père, époux [mari], garçon). Le tableau suivant présente les résultats :

	F	FSL	FWF
filles	5	3	2
garçon	0	0	0
mère	8	3	5
père	0	0	0
épouse	4	2	2
époux	2	1	1

Il ressort de ce tableau que la fréquence des cohyponymes de « femme » est toujours supérieure à celle des cohyponymes d'« homme », représentés seulement pour un mot sur trois (époux). Cette fréquence est même nulle pour « garçon » et « père ». Formellement donc, le silence partiel porte sur ces deux mots. Leur effacement crée un « trou lexical discursif » faisant apparaître le DOP comme un discours asymétrique où la procréation semble impliquer la fille et la mère (la femme), à l'exclusion du garçon et du père (l'homme). De quoi cette asymétrie est-elle le signe ? C'est à cette question que va tenter de répondre la réflexion qui suit.

4. ANALYSE

Nous examinerons le silence total avant de nous pencher sur le silence partiel.

4.1. Silence total

L'apparition tardive du DOP dans l'interdiscours, même après l'abrogation de la loi de 1920 en 1980, est essentiellement due à l'influence de l'islam confrérique qui évoquait des motifs religieux selon lesquels l'islam interdisait la propagande

antinataliste. Mais à partir de 1988, le nouveau courant islamique montra, avec des arguments islamiques jamais contestés, que l'islam n'interdisait pas la contraception. Samb (1988), Mbaye (1994) et le Réseau Islam et population, notamment, montrent que ni le Coran, ni la sunna et les hadiths, tradition et paroles du prophète de l'islam, n'interdisent la contraception ; elle est même encouragée, selon Samb (1988a), par les « voix autorisées », dont notamment l'Université Al Azhar du Caire. Aussi pense-t-il que la politique de planification familiale du Sénégal se heurte « plus au poids des traditions qu'aux enseignements de l'islam. » En renvoyant ainsi l'islam confrérique dans le champ de la tradition, l'auteur ne fait que reformuler la principale critique du nouveau courant islamique à l'islam confrérique. Toujours est-il que celui-ci garde le silence sur la question de la contraception.

Pendant son silence (1988-1997) dont nous parlions tantôt, la ligne rédactionnelle de WF connaît une évolution notable. À partir de 1993, l'objectif n'est plus d'être un quotidien islamique mais une presse libre œuvrant pour la démocratie. On pourrait donc penser que la religion n'explique pas, au moins à partir de 1993, le silence du journal ; d'autant plus qu'en 1988 déjà, le nouveau courant islamique avait apporté sa caution à la politique de population de l'État, en se fondant sur un argumentaire religieux. Et WF est l'un des initiateurs de ce courant. Malgré tout le motif religieux n'est pas absent de l'attitude de WF de 1993 à 1997. En effet le premier article du journal sur le DOP, une enquête en Iran du directeur de publication (Niass 1997), semble indiquer qu'il était à la recherche d'une caution religieuse spécifique, celle du chiisme iranien, comme le montre ce passage :

« ce qui est tabou chez le religieux au Sénégal semble tout à fait normal chez son alter ego d'autres lieux, surtout en Iran islamique [...] Et c'est d'autant plus remarquable qu'au début de la révolution islamique, l'option était pour une famille nombreuse pour répondre aux enseignements islamiques clairement expliqués par le prophète de l'islam [...] »

Ainsi, cette question ne doit pas être taboue au Sénégal où les chefs religieux ainsi que les autres acteurs devraient animer le débat sur la démographie.

WF aurait donc pour seule référence le modèle iranien. Une coïncidence, peut-être fortuite, mais certainement troublante semble accréditer cette idée : le changement de ligne rédactionnelle du journal a lieu la même année où, en Iran, toujours selon le même article, « Ce débat public a abouti à une loi votée en 1993 donnant une sanction religieuse au contrôle des naissances ».

Mais la thèse de l'influence iranienne résiste difficilement si l'on tient compte du fait que la politique de contrôle de naissance de l'Iran date de 1989 (Microsoft 2003), et que WF en était certainement informé. En réalité, en tant que formation discursive, cet organe semble écartelé entre l'islam confrérique, dont il est très proche, et le nouveau courant islamique, dont il est l'un des initiateurs. De ce point de vue, son silence est assimilable à celui de l'islam confrérique : seul l'attachement à une certaine conception de la tradition semble l'expliquer. L'article jouerait alors deux fonctions : il participe d'une part du processus de reformulation de l'interdiscours de l'islam confrérique, en témoigne la répétition de « religieux »

dans « le religieux au Sénégal » et « les chefs religieux ». Au Sénégal en effet, seul le guide spirituel d'une confrérie musulmane bénéficie de ce titre ; d'autre part, l'effort d'assimilation d'un composant extérieur se double, à un autre niveau, celui du journal en tant qu'organe de presse, d'un travail qui relève de la stratégie de communication consistant à marquer sa personnalité par la référence au chiisme¹¹ car dans le domaine de la presse où les contenus sont si peu différents – WF répète ce que le nouveau courant islamique dit depuis dix ans — un organe marque son identité en recourant à une stratégie discursive qui lui est spécifique. C'est une manière d'établir une certaine relation avec le lectorat : l'énonciation joue ainsi la fonction sémiologique de signe de reconnaissance, voire de ralliement. Le silence de WF est donc, en dernière analyse, une tentative d'utilisation de son influence en vue de renforcer la position de l'islam confrérique.

À la différence de l'islam confrérique et du WF, le silence de la tradition n'est pas le résultat d'une transaction. Il lui a été imposé par les autres formations discursives qui l'assimilent au paganisme. Cela revient à dire, dans le contexte du Sénégal, que le discours de la tradition n'a pas droit à l'existence. Ce point de vue est développé pendant les dix premières années du DOP, quand la formation discursive étatique et le nouveau courant islamique, prenant leurs représentations pour des évidences, croyaient que l'islam déterminait l'attitude des Sénégalais en matière de procréation. Lo (1988 : 15), dont le point de vue est largement représentatif, parlait ainsi de « coutumes » à « exorciser », de « traditions qui résistent aux assauts du modernisme ».

Cet ostracisme cachait mal cependant la forte influence du discours de la tradition dans l'interdiscours des formations discursives qui l'excluaient. Le discours de la tradition apparaît dans le DOP après dix ans de silence, lorsque les résultats de la politique de population se sont révélés insuffisants. Devant l'échec, le champ politique, avec le consentement tacite du nouveau courant islamique, se souvenant subitement des vertus du discours de la tradition, lui accorde un intérêt à la mesure du silence qui lui était imposé naguère. La réhabilitation s'opère en décembre 1997 et se traduit formellement par l'invention du concept de « communicateur traditionnel », sans doute pour moderniser la condition et la fonction de griot. Cette néologie consacre, autant qu'on peut en juger par les mots de la ministre du Développement social, la naissance du griot à un monde nouveau, celui des :

« partenaires au développement [qui] sont aujourd'hui prêts à collaborer avec les communicateurs traditionnels pour la diffusion de leurs programmes. Ce choix s'explique, selon elle, par le fait que les partenaires au développement [...] reconnaissent que les communicateurs traditionnels savent comment mobiliser les foules, et influencer positivement sur les comportements » (Mendy : SL 24/12/97).

Le silence de la tradition d'une part, celui de WF et de l'islam confrérique d'autre part, semblent différents, mais dans les deux cas, c'est l'autorité du savoir — ou ce qui est considéré comme tel — qui en impose. Tout simplement, le rapport de force

¹¹ L'argumentaire du nouveau courant islamique se fonde essentiellement sur le sunnisme.

empêche une existence autonome du discours de la tradition. Mais l'absence d'identité formelle ne semble pas un handicap insurmontable pour une formation discursive parce que celle-ci se définit avant tout à « partir de son interdiscours et non l'inverse. » (Maingueneau 1987 : 82). En l'occurrence, le discours de la tradition, par son influence dans l'interdiscours, transforme les autres formations discursives plus qu'elles ne le modèlent : il leur sert de référentiel. Ceci est tellement vrai que l'influence des courants religieux se mesure à leur capacité à assimiler les composants du discours de la tradition. Ainsi s'explique, au moins en partie, le caractère marginal du nouveau courant islamique.

4.2. Le silence partiel

Contrairement au silence total qui fait l'objet d'une relation conflictuelle entre les formations discursives (pour ou contre le DOP), le silence partiel portant sur « père » et « garçon » est le résultat d'un large consensus. Cette affirmation est autorisée par le matériau dont l'analyse a permis de repérer les occurrences de silence partiel. Il s'agit d'un échantillon représentatif du *corpus* global, qui comprend donc toutes les formations discursives, y compris l'islam confrérique pourtant formellement absent du *corpus* : c'est que WF constitue pour lui une sorte de cheval de Troie plus ou moins tacite dans le DOP.

La relation d'alliance générale dont procède le silence partiel ne doit pas cependant en cacher le côté paradoxal. Nous avons dit plus haut que le silence était un instrument de pouvoir, et comme tout pouvoir, c'est le dominé qui le subit. Le Sénégal étant, comme le souligne fort justement Benga (1997) « un pays où la culture et les traditions assignent aux hommes tous les pouvoirs », comment expliquer le silence sur « garçon » et « père » ?

Si, au-delà du formalisme de l'analyse lexicométrique, l'on tient compte, d'une part, de ce que nous avons appelé le demi-silence partiel sur « homme », d'autre part de la nature du discours à analyser, un discours sur la procréation, on peut raisonnablement conclure que le silence partiel dans le DOP porte sur la fonction procréatrice de l'homme, fonction dont l'expression lexicale est le mot « père ». Il se trouve que dans la société sénégalaise, le statut de l'homme-père coïncide avec celui de l'énonciateur légitime, c'est-à-dire, selon Foucauld (1969 : 68), celui qui a « le droit [...] traditionnel [...] spontanément accepté de proférer » le DOP. Ce père-énonciateur légitime est non seulement éprouvé par la crise économique, mais il est en plus confronté, sur le plan symbolique, à du savoir, des arguments difficile à réfuter.

Toutefois ces deux forces ne sont pas suffisamment puissantes, appropriées devrait-on dire, pour provoquer une mutation des valeurs traditionnelles attachées à la procréation – valeurs construites au fil du temps pour la conservation de la collectivité, donc éprouvées par la durée, et transmises de génération en génération comme un bien collectif sacré. Le problème posé par le DOP est donc celui de la relation entre la réalité objective et les représentations collectives dans lesquelles se pense la société sénégalaise. L'énonciateur légitime est conscient de ce conflit entre ses aspirations à la « modernité » et la force qui l'empêche d'y accéder. Il ne

manque pas de le répéter dans un langage parfois à l'allure martiale, avec une foi de néophyte dans ce qu'il appelle le « modernisme ».

À la pression exercée par la tradition, s'oppose celle de la formation interétatique, le FNUAP en l'occurrence. Prises en tenailles entre ces deux forces irrésistibles, les autres formations, WF et l'islam confrérique, mais aussi SL et le nouveau courant islamique, jouent de stratégie dans l'interdiscours pour une large alliance dont le ciment consiste en l'effacement de certains de leurs composants au profit de composants des formations les plus résolues. Cette recomposition de l'interdiscours consiste en une concession à la formation interétatique se traduisant par l'existence formelle du DOP, mais un DOP vidé de sa substance par l'effacement du père-énonciateur légitime qui détient tous les pouvoirs. Pour sauver de ces pouvoirs ce qui peut l'être encore, les formations intermédiaires réduisent le DOP en un discours de l'homme sur la femme, discours selon lequel le taux élevé de croissance démographique s'explique par le

fait que dans toute la sous-région sahélienne les femmes ont une vie féconde beaucoup plus longue par rapport à la moyenne mondiale. Elles commencent à concevoir leurs enfants dès le jeune âge (12-13 ans) et elles continuent jusqu'au-delà de quarante ans (Benga, *op. cit.*).

Le père-énonciateur légitime atteint ainsi le dernier degré de l'auto-effacement en laissant entendre que les femmes conçoivent seules des enfants dans « un pays où la culture et les traditions assignent aux hommes tous les pouvoirs. »

5. CONCLUSION

Notre objectif était de contribuer à l'étude discursive du silence comme phénomène langagier spécifique. Dans cette perspective, nous avons procédé à la distinction entre silence actualisé et silence réalisé. Le premier consiste en une suppression de signifiants recouvrables par les règles du langage verbal, tandis que le silence objet de notre réflexion relève du discours. Le silence réalisé une fois délimité, nous nous sommes attaché à montrer ses modalités que sont le silence total et le silence partiel. Pour éprouver son statut discursif, nous avons examiné son fonctionnement dans le DOP, en utilisant des concepts et des méthodes couramment employés en analyse du discours. Les résultats obtenus tendent à confirmer l'existence d'un statut discursif du silence. Mais la rareté des études systématiques sur cette question constitue une difficulté que la recherche devrait permettre de dépasser.

BIBLIOGRAPHIE

- Amossy (R.), 2000 : *L'Argumentation dans le discours. Discours politique, Littérature d'idées, fiction*, Paris, Nathan/HER.
- Benga (M.-L.), 1997 : « Les multiples enjeux de la santé de la reproduction », dans *Le Soleil*, 29/5/1997.
- Bourdieu (P.), 1998 : « La lutte féminine au cœur des combats politiques. De la domination masculine », dans *Le Monde diplomatique*, août 1998, p. 24.

- Bourdieu (P.), 1996 : « La télévision peut-elle critiquer la télévision ? Analyse d'un passage à l'antenne », dans *Le Monde diplomatique*, avril 1996, p. 25.
- Foucault (M.), 1969 : *L'Archéologie du savoir*, Paris, Gallimard.
- Lo (S. F.), 1988 : « Un préalable pour maîtriser le développement », dans *Le Soleil* 26/1/88, p. 15.
- Maigneueau (D.), 1987 : *Nouvelles Tendances en analyse du discours*, Paris, Hachette.
- Niass (S. L.), 1997 : « Iran, croissance démographique : les mollahs se “ préservatisent ” », dans *Walfadjri*, 6-7/6/1997, p. 9.
- Microsoft, 2003 : « Iran », dans *Encarta*, CD-Rom.
- Pulcinelli (E.) : 1991, « Hétérogénéités et silences », dans Parret (H.), édit., *Le Sens et ses hétérogénéités*, Paris, Centre national de la recherche scientifique. pp. 203-215.
- Samb (B.), 1988 : « Islam et planification familiale : la régulation des naissances, une nécessité dans le Tiers-monde » dans *Le Soleil*, 6/6/1988, pp. 8-9.
- Waquet (F.), 1998 : *Le Latin ou l'empire d'un signe. XVI^e – XX^e siècle*, Paris, Albin Michel.

TEXTE, TERME ET CONTEXTE

Rute Costa

centro de linguística da Universidade Nova de Lisboa — CLUNL

Avenida de Berna, 26-C

1069-061 Lisboa

m.rutecosta@mail.telepac.pt

***Résumé** : Pour que le texte soit compris comme un tout, il est nécessaire de prendre en compte des paramètres linguistiques et extralinguistiques qui interviennent dans tout acte de rédaction et de construction du sens. Le concept de contexte peut ainsi assumer diverses interprétations et peut déclencher toutes sortes d'ambiguïtés. En communication de spécialité écrite, le contexte extralinguistique exerce de l'influence au niveau de la construction du cadre communicatif, modulant ainsi les conditions de productions textuelles, facteur qui se reflète dans l'organisation des connaissances.*

En science terminologique, la relation entre l'extralinguistique – dans son sens le plus global – et le linguistique est une constante, étant donné que l'organisation et la hiérarchisation des connaissances sont primordiales pour une utilisation adéquate du terme en discours.

***Mots-clés** : texte de spécialité, type de discours, genre de discours, contexte linguistique, contexte extralinguistique, terme.*

1. INTRODUCTION

Réfléchir sur le statut du texte est de plus en plus d'actualité entre ceux qui ont des inquiétudes concernant les théories et les méthodologies en terminologie, car le texte est indubitablement un objet auquel l'on a recours dans quasiment tout travail de recherche en terminologie.

Que le but de la recherche terminologique soit la modélisation des connaissances ou encore la description et l'organisation des éléments qui permettent la construction du sens, le texte, en tant qu'objet d'analyse, joue un rôle primordial.

L'existence de grands *corpus*, l'augmentation de la capacité du matériel informatique, la mondialisation amènent les terminologues à se concentrer sur la gestion des données multilingues, à penser au développement des métadonnées afin de rendre l'intercommunication de spécialité plus efficace.

Si l'on envisage de travailler sur des *corpus* de spécialité, on se voit confronté à de nouvelles inquiétudes. Tous les *corpus* ne servent pas le même but, ce qui nous oblige à réfléchir sur les critères de composition de *corpus* et en conséquence à établir des critères de sélection de typologies de types et de genres de textes et/ou de discours. Car ce n'est qu'en fonction de l'établissement de ces critères qu'il est

possible de garantir l'adéquation des textes aux objectifs préétablis qui sont évidemment le garant de tout travail de recherche. Là encore une compétence additionnelle s'ajoute à celles déjà requises pour les terminologies ; statuer sur le texte de spécialité, en n'oubliant pas de réfléchir sur le statut des intervenants – auteur et locuteur – ainsi que sur le contexte de production et de réception.

Nous n'allons pas nous pencher sur l'importance et la pertinence de la différenciation entre texte et discours ; nous avons eu l'occasion de l'exposer ailleurs (Costa 2005). Toutefois, nous assumons pour notre propos, que « Le discours est conçu comme l'association d'un texte et de son contexte » (Maingueneau, 1996 : 29) et que, en accord avec Rastier, « [...] le texte est le lieu de rencontre entre le contexte et l'intertexte » (Rastier 1998 : 107).

De ce fait, nous croyons pertinent d'approcher le texte de spécialité – habitat privilégié des termes – à travers ses contextes sociodiscursifs et linguistiques, étant donné que tout acte de langage assemble une réalisation verbale et un propos social pour transmettre et accéder à la connaissance.

2. CONTEXTE LINGUISTIQUE ET CONTEXTE EXTRALINGUISTIQUE

Ainsi, le texte de spécialité peut, simultanément, être compris comme la production et le produit d'une communauté de communication restreinte (Maingueneau 1992) ou nucléaire (Beacco 1995). Dans le texte se concentrent tous les éléments linguistiques et extralinguistiques qui résultent de l'interaction du langage avec la vie sociale, ce qui fait que le texte peut être analysé en même temps comme un processus et comme un résultat.

L'importance des paramètres extralinguistiques, perceptibles dans l'analyse du texte en tant que produit d'une activité intellectuelle et sociale d'un individu ou d'un groupe d'individus, est au moment de la production textuelle un fait indéniable.

Diverses sont les désignations proposées pour se référer de façon non ambiguë à la réalité qui n'appartient pas à la sphère de la linguistique, mais qui appartient aux sphères sociales, psychologiques, pragmatiques et/ou physiques et qui interfère dans la production textuelle en ayant des conséquences sur le produit final, le texte.

A priori, nous désignerons comme *contexte* les réalités, linguistiques ou non, qui entourent le texte et sa production. Rastier introduit le concept de contexte comme élément indispensable à l'interprétation d'un texte. Pour lui, le contexte peut être compris de diverses façons : « [...] il est utilisé soit pour réintroduire une forme limitée de globalité dans l'interprétation, soit pour lier la langue à une extériorité en conditionnant la construction du sens à la connaissance de cette extériorité » (Rastier 1998 : 98).

De cette façon, le contexte peut assumer plusieurs significations, pouvant évidemment gérer des ambiguïtés d'interprétation et par conséquent d'usages. Bronckart est un des auteurs qui évite l'utilisation du terme contexte, ne l'estimant pas assez ample. Il préfère recourir au concept d'« extra-langage », qu'il présente comme un

concept opératoire. Il le définit comme suit : « [...] l'ensemble théoriquement infini de toutes les entités « mondaines » en dehors du langage » (Bronckart 1995 : 26). Mais quand, par besoin, il se rapporte au contexte, il le fait en l'utilisant comme synonyme d'« extra-langage », en distinguant ce dernier du concept de contexte qui recouvre d'après lui, de façon inéquivoque, les questions qui sont du niveau de la linguistique : « [...] le concept de contexte s'adresse à l'extra-langage, et qu'il faut éviter de le confondre avec l'environnement linguistique d'un énoncé. Pour désigner ce dernier, nous utiliserons de concept de « co-texte » [...] » (Bronckart 1995 : 27).

Rastier, de son côté, distingue ces deux types de contextes en considérant que le contexte linguistique se définit comme « un voisinage local » (Rastier 1998 : 106), le contexte extra-verbal comme une situation.

Pour que le texte soit appréhendé comme un tout, il est nécessaire que l'on ait en compte tous les paramètres linguistiques et extra-verbaux qui interviennent dans l'acte de la « parole ».

Dans son ouvrage, publié en 1997, Bronckart utilise l'expression « situation d'action langagière » pour se référer aux paramètres qui ont de l'influence sur l'analyse du texte, entendu comme un objet référentiel. Il nous dit que « Cette expression générale désigne les propriétés des mondes formels (physiques, sociaux et subjectifs) qui sont susceptibles d'exercer une influence sur la production textuelle » (Bronckart 1997 : 93). De tels paramètres peuvent être sous-divisés en « situation d'action langagière externe » et « situation d'action langagière interne ». La première concerne les caractéristiques du monde formel, c'est-à-dire le monde observable, la deuxième se rapporte à l'action intériorisée des individus, tant de l'auteur que du destinataire.

Pour se référer aux mêmes réalités, Desclés utilise une terminologie distincte. Il distingue les contextes externes, qu'il décrit comme faisant « [...] appel à des connaissances sur le domaine externe, sur les interlocuteurs, sur les conditions de dialogue, sur les buts poursuivis et les tâches à résoudre... », des contextes internes qui « [...] reposent sur des connaissances linguistiques qui entrent dans le co-texte d'une unité linguistique » (Desclés 1997 : 216). Alors que Bronckart utilise le terme *co-texte* pour désigner l'environnement linguistique d'un énoncé, Desclés l'utilise pour désigner l'environnement linguistique qui entoure l'unité linguistique en analyse. Pour cet auteur, le concept de co-texte englobe une réalité linguistique délimitée et courte, en comparaison avec le concept de co-texte de Bronckart, qui renvoie à une réalité linguistique qui peut être aussi vaste que peut l'être l'énoncé.

Rastier synthétise, de façon claire, les deux perspectives de contextes linguistiques présentes chez Bronckart et Desclés : « [...] on peut opposer deux conceptions du contexte linguistique : comme zone d'extension, relativement au signe ; ou de restriction, relativement au texte » (Rastier 1998 : 99). Le contexte linguistique remet à deux réalités différentes. D'un côté, le contexte qui entoure l'unité de signification en analyse, qui correspond à la notion de contexte utilisée par Desclés ; de

l'autre, le contexte vu comme une zone linguistique plus élargie, le texte dans son amplitude qui est en consonance avec le point de vue de Bronckart.

Le contexte linguistique, entendu comme une zone d'extension, permet l'élargir l'information conceptuelle, sémantique et syntactique de l'unité de signification à travers son analyse. Le contexte linguistique, compris comme une zone de restrictions, se définit comme une zone de localité, étant donné qu'« [...] un contexte n'est plus le contexte d'un mot, mais un passage du texte. Un mot peut certes servir à choisir un passage dans une recherche [...] ; mais ensuite, c'est le passage qui devient l'unité étudiée » (Rastier 1998 : 99).

Bronckart et Desclés ont pour objet d'étude le texte. Alors que Bronckart développe une méthode d'analyse basée sur un modèle psychologique, Desclés, lui, développe une méthode d'exploration contextuelle qui a pour objectif la construction d'un système basé sur des connaissances exclusivement linguistiques présentes dans le texte (*cf.* Desclés 1997 : 215).

Bronckart (1997) analyse le fonctionnement du texte et son architecture interne, en partant de principes théoriques sociodiscursifs sous-jacents à sa création. L'approche de Desclés (1997) est purement linguistique ; celui-ci prétend créer un modèle qui parte d'une description linguistique fine, parce qu'il considère que les unités linguistiques sont des indices pertinents pour le développement de tâches spécifiques. Ainsi, isoler et observer le contexte qui entoure une unité terminologique est primordial, une fois que celle-ci contient les éléments nécessaires pour mieux comprendre l'unité linguistique en analyse, en laissant des traces tout autour, qui sont passibles d'être identifiées.

Nous pensons que les contextes linguistiques et les contextes extralinguistiques se sous-entendent, dans la mesure où l'individu et l'objet se produisent réciproquement. Le contexte assume une fonction médiatrice entre l'individu et le monde objectif, idée que Rastier exprime de la façon suivante : « [...], de même que l'environnement est un concept intermédiaire entre celui du monde propre du sujet et celui du monde objectif, il semble que le concept de « contexte » soit un pseudo-concept, qui tout à la fois souligne et voile l'unité entre le mot et le texte, comme le lien entre la situation et la pratique en cours » (Rastier 1998 : 110).

Le contexte est un concept opératoire, médiateur entre l'intralinguistique, le linguistique et l'extralinguistique, où les intervenants sont les principaux protagonistes de la médiation. C'est pourquoi Feldman (2002) nous parle d'une communauté interprétative.

Prendre en compte le niveau extralinguistique va au-delà de la reconnaissance des concepts et du système conceptuel dans lequel le contexte s'insère ; cela va aussi dans le sens de l'appréhension des conditions de la mise en discours de ces connaissances et des conditions de l'interprétation.

3. INTERVENANTS

Dans le cas de la réalisation de la communication à travers le texte entendu comme un produit matériel avec des contours physiques finis, qui fixe un discours proféré par un ou plusieurs individus insérés dans des contextes spécifiques, la relation entre l'auteur et le lecteur se caractérise par le fait que, en général, la communication se réalise en l'absence d'un des intervenants, le texte étant le lien entre eux.

L'actualisation en discours est une activité coopérative qui a des implications au niveau de ses intervenants. Pour Charaudeau, « Tout acte de communication est un objet d'échange entre deux instances, l'une d'énonciation, l'autre de réception, dont le sens dépend de la relation d'intentionnalité qui s'instaure entre celles-ci » (Charaudeau 1997 : 15), ce qui présuppose trois lieux distincts associés à la production, à la réception de la transaction verbale médiatisée par le texte. Ainsi, la production s'intègre dans le lieu des conditions de production ; le produit, dans le lieu de la construction du discours ; et la réception, dans le lieu de l'interprétation.

D'après Charaudeau, le processus de transaction verbale se réalise d'après quatre principes qui se parfont : le principe de l'altérité, le principe de la pertinence, de principe de l'influence et le principe de la régulation (*cf.* Charaudeau 1995 : 98-100).

Pour que la communication se réalise, il est nécessaire que les intervenants aient des référents en commun et qu'ils soient en mesure de reconnaître l'univers de références de l'autre (principe de l'altérité), malgré le fait que l'auteur et le lecteur se reconnaissent comme des sujets structurellement et psychologiquement différents (principe de la pertinence). L'auteur produit un discours ou un texte avec l'intention d'exercer de l'influence sur l'autre, de le faire agir, de le modifier (principe de l'influence). Le lecteur, de son côté, consciemment ou inconsciemment, régule le jeu des influences de façon à viabiliser l'échange verbal (principe de la régulation).

Les intervenants sont une des composantes qui, d'après Bronckart, jouent un rôle dans les paramètres extra-verbaux, plus spécifiquement dans la « situation d'action langagière interne » que nous avons identifiée comme étant l'action intériorisée des individus, tant de l'auteur que du destinataire.

Cet auteur qualifie les intervenants, les uns d'actifs, les auteurs, les autres de passifs, les destinataires. L'intervenant actif est le producteur du texte qui incorpore toute situation sociale et psychologique. L'intervenant passif est le destinataire du texte, le lecteur qui tout comme l'intervenant actif incorpore une situation sociale et psychologique, mais dont la fonction est de lire, de comprendre et d'interpréter le texte.

L'intervenant actif, dans ce cas, le spécialiste producteur d'un texte de spécialité, structure son savoir en fonction d'un public modèle, duquel il possède une image prototypique. Le public est ainsi constitué par un ensemble d'individus qui, dans un premier moment, sont considérés comme passifs, dans la mesure où, au moment où le texte est construit, ils n'exercent aucun type d'action directe sur le texte.

Ils sont néanmoins considérés comme intervenants, car ils sont présents dans l'esprit de l'auteur qui construit une image du lecteur potentiel, que Eco désigne comme lecteur-modèle (1979). C'est sur la base de l'image que l'auteur a de la compétence de son lecteur-modèle que le spécialiste structure une complexité conceptuelle et linguistique qui se reflète dans le texte de spécialité. La compétence de haut niveau, attribuée au lecteur-modèle, permet au spécialiste d'élaborer un texte avec un niveau de scientificité très élevé qui, généralement, se répercute dans un texte dense et rigoureux, incompréhensible pour le non-spécialiste qui ne connaît pas les concepts, les réseaux conceptuels et référentiels qui composent le texte.

Tout texte a pour fonction d'être lu : « [...] lire, c'est en toute hypothèse, enchaîner un discours nouveau au discours du texte » (Ricœur 1986 : 152). Le lecteur intervient au moment où l'auteur du texte considère son travail de rédaction fini et le met à disposition. Le changement des entourages historiques et sociaux dans lequel le texte a été produit a des conséquences au niveau des connaissances véhiculées par le texte : « Si l'entour change, le contenu du texte change aussi puisqu'il est immanent à une situation de communication maintenant modifiée. En règle générale, dans le cas d'un changement d'époque ou de culture, il s'appauvrit, par déperdition des connaissances » (Rastier 1989 : 51).

Dans ces circonstances, le lecteur assume le rôle de lecteur actif, avec des fonctions et des degrés d'intervention différents de celui de l'auteur. D'après Rastier, le lecteur assume deux attitudes différentes : celle de la lecture productive et celle de la lecture descriptive. Dans le cas de la lecture productive, le lecteur réinterprète le texte alors que dans le cas de la lecture descriptive, son objectif est de restituer le contenu du texte.

Nous sommes convaincue que le lecteur spécialiste fait, généralement, une lecture productive. Le lecteur spécialiste s'approprie du texte, restructure l'information reçue à la lumière de ses propres savoirs et connaissances. En lisant le texte, le lecteur spécialiste ajoute un atout à son savoir, ajoute du savoir au savoir qu'il détient déjà et, théoriquement, lecture après lecture de ce texte et d'autres textes, il acquiert des compétences pour produire de nouveaux textes. Les textes par lui produits sont ajoutés aux textes déjà existants et, conjointement, il provoque des processus semblables chez d'autres lecteurs. Comme le dit Ricœur : « Dans cette chaîne, les premiers interprétants servent de tradition pour les derniers interprétants [...] » (Ricœur 1986 : 159).

4. LE TERME EN CONTEXTE

Les termes ont la particularité d'être des unités lexicales de spécialité, par le fait de représenter des connaissances spécifiques à un domaine du savoir, reconnues et partagées par les membres d'une communauté de spécialistes. Cette connaissance est le résultat de la maîtrise d'un savoir qui se traduit par la connaissance que l'on possède des relations établies entre les termes formant une structure lexicale et qui dans un contexte textuel et/ou discursif dénomment les relations entre concepts,

dans un système conceptuel donné. Rey considérait le nom, au sens logico-philosophique de l'anglais *name*, « [...] comme l'objet même de la terminologie : en effet, un nom définissable à l'intérieur d'un système cohérent, énumératif et/ou structuré, est un terme ; le contenu de sa définition correspond à une notion (concept), analysable en compréhension » (Rey 1979 : 22).

De toute évidence, du point de vue de la communication, le terme se comporte au niveau syntaxique, morphologique et lexical, en langue et en discours, comme toute unité lexicale. Cette idée, Kocourek la manifeste comme suit : « Les termes et les non-termes du fond lexical entier d'une langue ont en commun l'appartenance à cette langue, d'où résulte leur dépendance au système phonologique, morphologique et lexical de cette langue, leur subordination, dans le contexte, aux lois syntaxiques et aux tendances sémantiques et, enfin, leur validité inégale dans le temps, dans l'espace et la réalité sociale » (Kocourek 2001 : 296).

L'unité terminologique a, pendant de longues années, été le principal objet d'étude des terminologues. Mais, est-ce encore toujours vrai ? L'unité terminologique est un axe obligatoire pour chaque terminologue, dans la mesure où l'unité terminologique est dénomminative et la dénomination est l'objet même de la terminologie.

Mais l'unité terminologique ne bénéficie plus d'une attention exclusive. Le terminologue se concentre de plus en plus sur l'établissement du rapport, du lien ou encore de la relation entre deux ou plusieurs concepts au sein d'un même système conceptuel. Cette perspective peut être corroborée par une affirmation de Rey, qui préconise que : « [...] la terminologie étudie des signes. Ces signes se manifestent au moyen des formes des langues naturelles (mots, etc.), leur rapport avec ses formes doit être précisé » (Rey 1979 : 19).

Or, c'est dans cette notion de *rapport* que se situe, me semble-t-il, une partie importante du débat, car c'est dans le rapport entre ce qui est dénommé et la dénomination que se trouve l'essence du travail en terminologie et qui en définitive se trouve au cœur des travaux actuellement en cours. Ici, la relation entre le contexte extralinguistique et le contexte linguistique nous semble une évidence.

En contexte linguistique, des séquences linguistiques non dénomminatives, tels que les collocations et les phraséologies, prennent de l'ampleur dans les études terminologiques. Souvent, celles-ci sont appréhendées comme des microcontextes linguistiques privilégiés pour comprendre le comportement morphosyntaxique et sémantique de l'unité terminologique. C'est à partir de l'observation et de la description du comportement du terme dans ce que nous pouvons appeler son *micro-contexte linguistique*, que nous sommes en mesure d'élaborer des règles linguistiques et, postérieurement, des règles formelles pour le traitement automatique des langues.

5. NOTES CONCLUSIVES

S'il est vrai que l'organisation des connaissances qui se trouvent à un niveau extralinguistique est un des objectifs primordiaux de la recherche en terminologie, il

n'est pas moins vrai que c'est à travers l'acte de la parole, c'est-à-dire le discours, que nous pouvons accéder à la représentation de l'organisation des connaissances. La parole étant un moyen privilégié de représentation du monde et l'endroit de construction du sens ; mais elle n'est pas le monde en soi. La difficulté de la théorisation réside justement dans le fait que ces deux réalités – le monde et sa représentation discursive – forment une association durable et réciproquement profitable, qui nous amène à citer Aubenque, qui dit : « C'est parce que les choses ont une essence que les mots ont un sens » (Aubenque 1962 : 128).

Si d'un côté, il y a symbiose, de l'autre la distinction théorique est indispensable pour comprendre la différence entre la perspective linguistique, plutôt sémasiologique et la perspective conceptuelle, plutôt onomasiologique, qui ne s'excluent pas mutuellement, mais se complètent, ceci étant finalement de la compétence du terminologue. Cette opposition presque dichotomique nous semble aujourd'hui bien réductrice. En effet, le recours aux grands *corpus* et, en conséquence, le développement des linguistiques de *corpus*, nous font reconsidérer les méthodologies utilisées pour l'extraction des termes, le traitement de l'information, et aussi pour la construction des ontologies.

En conclusion, nous pensons que pour appréhender le terme dans ses multiples facettes, il faut l'observer dans son contexte extralinguistique et dans son contexte linguistique.

Obtenir de bons résultats à partir du traitement automatique de la langue sur *corpus* présuppose une différenciation théorique claire entre la représentation des connaissances et la construction du sens en discours. La première dérive de l'observation de l'état des choses, la deuxième rend compte de la diversité dénominative inhérente à une communauté linguistique spécialisée.

En ce qui concerne le niveau linguistique, nous identifions le niveau du texte qui est le résultat d'une activité sociale qui se manifeste dans le type et le genre de discours : « Les régimes interprétatifs des types de discours (politiques, religieux, etc.) sont enfin spécifiés par les contrats interprétatifs propres aux genres » (Rastier 1998 : 109). Et c'est dans le texte, entendu comme une occurrence du discours, que nous abordons le contexte et le co-texte qui nous permet une analyse lexicale, morphologique, syntaxique et sémantique épurée.

BIBLIOGRAPHIE

- Aubenque (P.), 1962 : *Le Problème de l'être chez Aristote*, Paris, PUF.
- Bronckart (J. P.), 1997 : *Activité langagière, Textes et Discours. Pour un interactionisme socio-discursif*, Lausanne, Delachaux & Niestlé.
- Charaudeau (P), 1997 : *Les Discours d'information médiatique. La Construction du miroir social*, Paris, Nathan.
- Costa (R.), 2005 : « Pluralité des approches théoriques en Terminologie », dans *Proceedings of LSP 2005 – New Trends in Specialized Discourse*, Bergamo, Genève, Springer-Verlag [sous presse]

- Costa (R.), 2005 : « *Corpus de spécialité : une question de types ou de genres de textes ou de discours* », dans *Actes du colloque en hommage à Philippe Thoiron*, Lyon, PUL [sous presse].
- Costa (R.), 2001 : *Pressupostos teóricos e metodológicos para a extracção automática de termos multilexémicos*, thèse de doctorat, Universidade Nova de Lisboa, Lisboa.
- Costa (R.) et Silva (R.), 2005 : « An approach towards the linguistic dimension of Terminology », dans *Proceedings of CIT 2003 – The state of theories*, Lisboa, Colibri [sous presse].
- Costa (R.) et Silva (R.), 2004 : « The Verb in the Terminological Collocations. A Contribution to the Development of a Morphological Analyser. MorphoComp », dans Lino (T.), Xavier (F.), Ferreira (F.), Costa (R.) et Silva (R.), édit., dans *Proceedings of the IV International Conference on Language Resources and Evaluation – LREC 2004*, vol. IVth, Lisbon, ELRA, pp. 1531-1534.
- Desclés (J.-P.), 1997 : « Systèmes d’exploration contextuelle », dans *Co-texte et calcul du sens*, Caen, PUC, p. 215-231.
- Eco (U.), 1993 : *Leitura do Texto Literário. Lector in Fabula*, Lisboa, Presença.
- Feldman (C. F.), 2002 : « Les genres du discours comme modèles mentaux et culturels : l’interprétation dans une communauté culturelle », dans *Une Introduction aux sciences de la culture*, Paris, PUF, p. 215-228.
- Kocourek (R.), 2001 : « Le terme et sa définition », dans *Essais de linguistique française et anglaise. Mots et Termes, Sens et Textes*, Louvain-Paris, coll. « Bibliothèque de l’Information Grammaticale », Éditions Peeters, pp. 271-297.
- Rastier (F.), 1995 : « Le terme : entre ontologie et linguistique », dans *Banque des mots*, Paris, CILF, pp. 35 – 65.
- Rastier (F.), 1991 : *Sémantiques et Recherches cognitives*, Paris, PUF.
- Rastier (F.), 1989 : *Sens et Textualité*, Paris, Hachette.
- Ricœur (P.), 1986 : *Du texte à l’action. Essais d’herméneutique, II*, Seuil, coll. « Esprit », série « Essais », Paris.
- Mangueneau (D.), 1992 : « Le “tour” ethnolinguistique de l’analyse du discours », dans *Langages* 105, Larousse, Paris, pp. 114-125.
- Mangueneau (D.), 1998 : *Analyser les textes de communication*, Paris, Dunod.
- Thoiron (Ph.) et Béjoint (H.), 2005 : « Is terminology a question of terms ? », dans *Proceedings of CIT 2003 – The state of theories*, Lisboa, Colibri [sous presse].
- Silva (R.), Costa (R.), Ferreira (F.), 2004 : « Entre langue générale et langue de spécialité. Une question de collocations », dans *Étude de Linguistique Appliquée – ELA*, Didier Érudition, Paris, pp. 347-359.

TERME ET MICROCONTEXTE. LES PRÉDICTIONS SPÉCIALISÉES

Pierre Lerat

laboratoire de linguistique informatique
Université Paris XIII
93430 Villetaneuse

Résumé : *Les travaux sur les contextes non linguistiques et linguistiques ont fait avancer la terminologie, mais elle se heurte maintenant au non-corpus du Web, hétérogène, ouvert et instable. Le rapprochement des travaux ontologiques et terminologiques pose la question du calibrage des chaînes de caractères pertinentes, qui ne sont pas nécessairement des termes au sens classique. Une heuristique commune pourrait consister à privilégier comme formulations de connaissances spécialisées des jeux de prédicats (souvent d'ordre supérieur) et d'arguments.*

Mots-clés : *terme, ontologie, schéma d'arguments, prédication spécialisée.*

1. INTRODUCTION

La terminologie a d'abord été uniquement une onomasiologie. Il lui en est resté une forte composante paradigmatique, conceptuelle et normalisatrice. La conception du contexte qui va avec ces caractéristiques est celle du contexte situationnel. C'est lentement que progresse l'approche inverse, syntagmatique, linguistique et empirique. En particulier, le problème du calibrage des chaînes de caractères pertinentes est posé par la recherche libre d'informations spécialisées sur le Web. Le terme conçu comme « mot, syntagme, symbole ou formule désignant un concept propre à un domaine d'emploi » (Pavel et Nolet 2001 : 117), selon une tradition qui remonte à Wüster et à l'école russe (voir notamment Felber et Budin 1990 : 15), est relativisé en particulier par la variabilité des conceptualisations, par celle des domaines et par celle des expressions.

Pour limiter ma démonstration à un cas lexicalement et cognitivement bien circonscrit, j'ai choisi la question de la TVA réduite, qui est matière à débats dans l'Union européenne. Le moteur de recherche utilisé est Google, au moyen de requêtes sur des chaînes de caractères entre guillemets (qui ne distinguent pas TVA, T.V.A., t.v.a. et tva).

Le plan adopté est le suivant : la notion de « terme » est d'abord confrontée à celle de « contexte », puis à celle d'« ontologie » et enfin à celle de « prédication spécialisée ».

2. TERME ET CONTEXTE

2.1. Le terme hors contexte

La notion de terme « désignant une notion » est la plus classique, et elle prévaut encore dans les milieux de la normalisation, notamment industrielle. Elle est concurrencée par une approche plus linguistique, où le terme est vu comme une « unité lexicale définie dans les textes de spécialité » (Kocourek 1991 : 180). Une version minimaliste en est l'idée que « le terme est une variété fonctionnelle du nom commun » (Sager 2000 : 53). Il existe désormais également une version maximaliste, celle où est considérée comme « candidat-terme » une séquence telle que *révéler une lésion osseuse* (Bourigault *et al.* 2004 : 105). L'idée est celle d'un terme long, on reste donc dans le cadre d'une approche paradigmatique ; vu syntagmatiquement, ce groupe de mots constituerait une prédication spécialisée, c'est-à-dire une expression prédicative et un argument technique. J'y reviendrai, car je vois là une conception du microcontexte à privilégier.

2.2. Le terme dans le texte

Quand on ne dispose pas d'une théorie de la phrase, on peut être tenté par un « substitut de contexte, très limité, aléatoire et non problématisé » : ce qui vaut pour la phraséologie selon Slodzian (2000 : 72) vaut aussi pour les cooccurrences et les collocations.

La cooccurrence a pour elle le respect des expressions linguistiques « telles quelles » (*ibid.* : 73). Elle doit son succès à celui de la linguistique textuelle, et aussi à des travaux de traducteurs, en particulier le livre de Cohen (1986). Il est remarquable que les « cooccurrents », dans ce livre, soient précisément les noms d'opérations et d'états, c'est-à-dire des prédicats verbaux, nominaux ou adjectivaux ; ainsi, la tendance de la bourse est haussière ou baissière, à la hausse ou à la baisse, elle monte ou descend etc.

La phraséologie est un problème de linguistique générale qui n'a pas de rapport avec les connaissances lexicalisées dans des termes ; elle est importante pour l'étude du figement lexical (voir Mejri 1997), mais elle est dépourvue de pertinence cognitive.

La collocation correspond à « une multiplicité de rapports » (Kocourek 1991 : 198). Elle a néanmoins été rendue productive par la terminologie computationnelle à date récente. L'idée d'associer les mots qui dans un même *corpus* partagent les mêmes collocations est loin d'être stérile ; ainsi, « le couple de contextes *administration-introduction* permet de dégager la classe des produits pharmaceutiques » (Le Moigno *et al.* 2002 et Bourigault *et al.* 2004 : 105). Il faut toutefois ajouter que la portée d'un tel résultat n'est pas transposable en dehors d'un *corpus* de traumatologie en réanimation chirurgicale ; il en irait autrement, par exemple, dans un *corpus* administratif. Ce que je retiens pour le moment, c'est que les collocations *administration* et *introduction* sont des noms prédicatifs suivis de leurs arguments appropriés. On est très loin du chaos auquel conduiraient de simples collocations

par catégorie grammaticale (la classe des noms inclut à la fois nominalisations et noms concrets, par exemple).

2.3. Le terme dans la proposition

Les termes ne servent pas seulement à désigner : ils sont porteurs de vérités. Les terminologies religieuses sont le vocabulaire de textes où les croyances sont supposées vraies, les terminologies grammaticales reflètent des consensus et des conflits sur la façon la plus exacte de décrire les faits de langue etc. Les termes ne sont pas exclusivement des noms de connaissances : ce sont aussi des éléments de discours spécialisés, qui participent à la logique de ces discours ; par exemple, *copier un fichier* n'est pas la même opération selon que l'on travaille à la main ou sur ordinateur, les procédures et les conséquences s'en ressentent.

On a un peu oublié cette dimension logique introduite par l'école de Vienne et présente chez Sager (1990 : 185). Certes, il s'agit d'une simple logique du premier ordre, mais les ontologies et les *WordNets* en tirent encore parti, et le formalisme des réseaux sémantiques n'est « qu'une représentation graphique de la logique des prédicats du premier ordre » (Rastier 2004 : 25).

L'idée que la phrase simple est analysable selon le modèle de la proposition élémentaire n'est pas neuve, et elle est notamment à la base de deux modèles souvent testés dans les traitements automatiques des langues. L'un est le modèle des fonctions lexicales au sens de Mel'čuk. Appliqué à une terminologie de ce qu'on appelle en France la grande distribution, il semble avoir posé des problèmes d'adéquation et de complétude (voir Dancette et L'Homme 2005 : 114). De toute façon, il impose le détour par une métalangue en latin, tout comme *WordNet* par une métalangue en anglais, et il est si complexe qu'il pose des problèmes d'encodage et de décodage.

Une théorie plus simple, universelle et portable, est celle des schémas d'arguments. On entend par là « un prédicat accompagné de ses arguments » (Gross et Mathieu-Colas 2001 : 70). C'est toutefois un modèle de phrase, non d'énoncé, du fait de sa généralité.

Quel rapport avec la terminologie ? On peut observer que l'usage d'un prédicat spécialisé, tel que *promulguer* ou *promulgation*, appelle celui d'arguments fortement contraints, tels que *loi* et *président de la République* ; s'il en va de même dans toute langue spécialisée, *mutatis mutandis*, alors il y a lieu de s'attacher aux schémas d'arguments spécialisés (voir Lerat 1988 : 18). La pertinence des résultats escomptables peut s'observer à propos de thématiques très variées, qui vont de la médecine (voir Gross et Mathieu-Colas 2001) au football (voir Gross et Guenther 2002).

Je propose maintenant d'adapter cette approche de trois façons. D'abord, en s'attachant à des prédications spécialisées (= une expression prédicative et un argument *in situ*) dans le discours (au vu de la variabilité que présentent les énoncés sur le Web) ; il y a là en effet un grand intérêt, d'une part parce que les expressions prédicatives spécialisées représentent les actions, états, processus et propriétés ca-

ractéristiques d'un métier, d'autre part parce qu'autour d'elles se distribuent plus ou moins les arguments typiques.

La deuxième suggestion est de prendre en compte les « prédicats en position d'argument » (Karolak 1989 : 48), comme *taxation* dans l'exemple suivant :

[1] « modification de la taxation des services » (1) / « taxation des services » (331)

C'est ce que fait la syntaxe sémantique dans la tradition de la linguistique polonaise : *taxation* y est ici un « argument prédicatif » (Bogacki 1992 : 72), et *modification* un « prédicateur dominant » (Nawakowska 1992 : 446).

La troisième est de ne pas négliger les arguments propositionnels, notamment après les verbes de modalités, comme ci-dessous :

[2] « Nous espérons que les livres numériques seront taxés à 5,5 % »

Il est en effet essentiel, dans la prise en compte des phrases complexes, d'élargir l'application du modèle aux prédicats d'ordre supérieur (voir Allwood *et al.* 1989 : 148). La logique classique parle en pareil cas de prédicats du second ordre, mais cette dénomination est devenue polysémique sous l'influence de Lyons (voir Rastier 2004 : 20), en sorte que « prédicat d'ordre supérieur » est préférable comme étant sans ambiguïté : dans une chaîne de prédications, « les prédicats d'ordre supérieur ouvrent des positions pour des structures sémantiques complexes constituées par d'autres prédicats » (Karolak 1989 : 15), et la relation est récursive.

Il existe des expressions prédicatives qui sont elles-mêmes techniques ; c'est le cas dans les 3 exemples suivants :

[3] « un client acquitte une TVA sur les services »

[4] « un prestataire de services reverse à l'État la TVA »

[5] « l'État prélève une TVA sur les services »

Il arrive aussi que seul l'argument soit technique, comme dans « payer la TVA ».

3. TERME ET ONTOLOGIE

3.1. Un rapprochement significatif

On parle depuis peu de « ressources terminologiques ou ontologiques » (Aussenac-Gilles et Condamines 2004, titre), ou même de « ressources termino-ontologiques » (Bourigault *et al.* 2004 : 92), voire de « termontographie » (Kerremans 2005, titre). Pourquoi ?

3.2. Les ontologies de domaines

Laissons de côté les ontologies générales, qui sont une modélisation lexicale du sens commun (voir Rastier 2004 : 38). Ce qui a un rapport plus direct avec la terminologie, c'est l'ontologie de domaine, c'est-à-dire « les concepts, relations et contraintes effectivement utilisés pour modéliser un domaine donné » (Zweigen-

baum 1999). La visée est un traitement automatique, mais la méthodologie est parente.

L'ancrage sur un domaine est une faiblesse commune aux ontologies et aux terminologies, du fait de la relativité des domaines par rapport à des « points de vue » (de Bessé 2000 : 187). En d'autres termes, « la définition d'un « domaine » est arbitraire et relative aux objectifs définis dans un projet terminologique » (L'Homme, Heid et Sager 2003 : 153). Par exemple, la TVA, si l'on en juge par les professions concernées, intéresse évidemment la fiscalité et la politique, mais aussi le bâtiment, l'édition, l'eau, l'électricité et le gaz, la gestion, l'immobilier, la restauration, etc.

La terminologie classique et l'ontologie partagent une autre faiblesse, qui est celle des thésaurus : la tendance à ne prendre en compte qu'un vocabulaire contrôlé. Ainsi, dans la présentation d'une ontologie juridique, on peut lire ceci :

[6] « Nous sommes intéressés par le terme *taxation*, mais non par le terme *taxation à 20 %* » (Lame 2003).

Dompage, on se prive ainsi d'accès à des contenus précis.

3.3. Une tradition commune : l'approche paradigmatique

L'arbre de domaine, qui se veut hiérarchie cognitive et hiérarchie lexicale à la fois, utilise en priorité la relation *is_a*, dans une métalangue issue de l'intelligence artificielle dans la langue de ses concepteurs. Ses formulations lexicales sont multiples en français, à commencer par *être un*, qui se décline aussi en *être une*. Immergées dans des documents, ces deux chaînes de caractères n'introduisent pas que des énoncés définitoires, et quand c'est le cas ils ne font pas apparaître forcément le même hyperonyme comme définisseur :

[7] « un taux est un » (20) –> calcul (prédicat d'action), pourcentage, rapport (prédicats d'état)

[8] « un taux est une » (10) –> comparaison, mesure (prédicats d'action)

[9] « une taxe est un » (38) –> prélèvement (prédicat d'action), impôt (objet)

[10] « une taxe est une » (59) –> contribution (prédicat d'action), somme d'argent (objet)

En outre l'héritage des propriétés, dont les problèmes sont bien connus dans le cas de la méronymie, ne va pas toujours de soi dans l'hyponymie non plus, parce que les lexiques ne sont pas des systèmes logiques mais des usages sociaux.

P. ex. : une propriété de la restauration en France, à cette date, est que la TVA y est de 19,6 % ; or on peut à la fois affirmer que la restauration rapide est un sous-ensemble de la restauration, professionnellement, et que fiscalement la TVA applicable y est de 5,5 %.

Une distinction classique en intelligence artificielle gagnerait à être maniée systématiquement en terminologie : celle entre classe et instance, c'est-à-dire « membre

de ». Il importe de distinguer une classe conceptuelle, comme celle des restaurants, une instance, comme Mac Donald's, et un objet référentiel, comme le Mac Donald's désigné dans tel discours.

Comme les ontologies de domaines sont de plus en plus dépendantes de la linguistique textuelle, leur approche « de bas en haut » les conduit à rencontrer comme collocations les expressions prédicatives, qui ne sont pas reconnaissables automatiquement, contrairement aux classes grammaticales, assez bien maîtrisées par les analyseurs. En particulier, ce que Grabar et Hamon (2004 : 74) appellent des « relations transversales » n'est rien d'autre que des expressions prédicatives à morphologie variable, comme ci-dessous :

[11] « est responsable de, a la responsabilité de »

[12] « a la responsabilité de, est en charge de »

La limite de ces « relations transversales » est leur pertinence restreinte à un texte ou à un *corpus*. Par exemple, le caractère causal d'*entraîne* dans « entraîne un surcoût » (*ibid.* : 72) est déterminé par le sujet et le complément : dans « Jean entraîne les juniors » il en irait tout autrement. Or, « il est important de noter la dépendance de l'ontologie du point de vue du concepteur » (Bourigault *et al.* 2004 : 91). Il est non moins important, et à méditer en terminologie même, que tout *corpus* est suspect de produire de simples effets de *corpus* ; c'est inévitable dès lors que « le *corpus* n'a d'existence et de validité que dans le cadre de l'expérience qui l'a suscité » (Slodzian 2000 : 74).

4. TERMES ET PRÉDICATIONS SPÉCIALISÉES

4.1. Terme et Web

L'utilisateur du Web est face à des éléments de discours variés, institutionnels et non institutionnels, scientifiques ou mercantiles etc. Il lui faut ne pas perdre pied car « pratiquement, dans un contexte comme la consultation du Web via un moteur de recherche, c'est lui qui réalise la tâche de recherche » (Aussenac-Gilles et Condamines 2004 : 81). Il est seul avec sa culture, et il lui faut ajuster sa propre catégorisation, « qui est le résultat d'une interaction entre le langage et l'esprit » (Temmerman 1997 : 55, n° 3), à la catégorisation opérée par l'auteur de chaque énoncé consulté.

Il ne saurait être question de traiter le Web comme un *corpus*, car c'est par nature un non-*corpus*, voire un anti-*corpus*. Il ne présente en effet aucune des caractéristiques attendues classiquement d'un *corpus* : il n'est ni homogène, ni clos, ni même stable. Il importe d'évaluer constamment les sources, faute de quoi on ne saurait faire du travail correct ; ainsi, le contenu de la séquence *TVA à taux réduit* varie d'un pays à l'autre : 5,5 % en France (sauf les DOM et les TOM), 6 en Belgique, 7 en Allemagne, 5 en Grande-Bretagne etc.

Comme je l'ai écrit dans un article récent (Lerat 2005), le salut ne peut venir, en l'occurrence, que d'une bonne formation philologique : critique des sources, validation par des signatures dignes de foi, comparaison des résultats.

Pour les sources, un filtrage rigoureux s'impose, tout le monde est d'accord sur ce point. Les textes les plus représentatifs de toute terminologie, à mon avis, appartiennent à l'une des catégories suivantes : textes normatifs et textes scientifiques (voir Sager 1990 : 159), mais aussi textes de professionnels représentatifs. Par exemple, une expression comme « hors du champ de la TVA » (86) est administrative, politique et professorale, tandis que « TVA à taux réduit » (4360) se trouve aussi sur les sites des groupements de professionnels.

4.2. Clés linguistiques d'accès à l'information spécialisée

Un premier principe, paradoxal par rapport à la tradition de la linguistique textuelle, est la recherche de fréquences relativement basses, car sinon il faut renoncer à se servir du Web :

[13] « taux réduit » (69 800) –> « taux réduit de TVA » (13 700) –> « taux de TVA réduit » (5 830) –> « TVA à taux réduit » (4 360) –> « TVA au taux réduit de 5,5 % » (149, donc un nombre de pages limité et rien que des contextes pertinents dès les premières pages : Sénat, artisanat, banques etc.)

Un autre principe concerne le calibrage des chaînes de caractères à interroger : comme l'important est le microcontexte des microcontextes, il paraît de bonne méthode de privilégier les chaînes de schémas d'arguments.

P. ex. : *TVA au taux réduit de 5,5 %* n'est ni un terme simple ni un nom composé, mais une chaîne de prédications : *TVA* (prédicat d'action en position d'argument) qui a un *taux* (prédicat de propriété en position d'argument par rapport à *réduit*) qui est de *5,5 %* (prédicat de propriété ayant pour argument *taux*) et qui est *réduit* (prédicat de propriété). On peut parler en pareil cas de « prédicats adjoints » (Karolak 1989 : 51).

Les prédications spécialisées ne sont pas les seules bonnes clés d'accès. Par exemple, le code documentaire « DOC/2008/2004-FR », sur Google, conduit directement à la réglementation communautaire concernant les taux de TVA. De même, « Cour de Justice internationale » (4 280), dénomination longue, est une voie évidente pour accéder à l'état actuel des connaissances sur cette institution, mais on améliore grandement la performance en allant jusqu'à « Cour de justice internationale de La Haye » (284). Ce qui est à éviter est la recherche à partir d'une formulation brève, car la matière est alors pléthorique.

Ce qui est vrai des noms communs l'est aussi des noms propres. Une expérience intéressante a été conduite sur *Leónidas*, qui relève du domaine de l'astronomie en espagnol (en français on dit *les Léonides* pour désigner ces étoiles filantes visibles vers la mi-novembre). Il se trouve que, à l'accent près, c'est également le nom d'un chocolatier belge et d'un héros grec. L'auteur (e) a réalisé un filtrage par langue (*Leónidas* en espagnol, *Leonids* en anglais), par domaine et par types de colloca-

tions, en sorte que les résultats obtenus sur un *corpus* spécialisé tiré du Web sont significatifs (voir Sánchez-Gijón 2005), mais que de travail ! On gagnerait certainement du temps et de l'argent en partant de ce qu'on sait des Léonides (et si l'on ignore tout du domaine ce n'est pas la peine de s'y mettre).

C'est tout le problème des langues spécialisées : dans la mesure où ce sont les vecteurs de connaissances spécialisées, il faut une bonne connaissance de la langue et un minimum de familiarité avec le domaine. Il faut aussi une bonne tactique, à défaut de recettes. En attendant mieux, je propose d'essayer les prédications spécialisées que suggère sa propre culture telle qu'elle est, aidée par les ressources accessibles, notamment les dictionnaires et encyclopédies.

5. CONCLUSION

En somme, il y a deux façons de traiter les microcontextes des termes. Classiquement, on part d'un terme prototypique, à la Sager, c'est-à-dire un nom commun simple ou complexe, selon la tradition des documentalistes et des commissions de terminologie, et on travaille sur ses cooccurrents ou collocations ; on débouche alors sur des expressions prédictives, comme le fait la terminologie computationnelle dans trois cas repérés ici (Bourigault *et al.*, Grabar et Hamon, et aussi Sánchez-Gijón, sous couvert de « relations non hiérarchiques »).

Je propose un cheminement inverse, au vu des tactiques de recherche qui réussissent sur le Web : partir d'une prédication spécialisée que l'on essaie, et bricoler linguistiquement et cognitivement autour ; on débouche alors sur des termes prototypiques qui ont de bonnes chances d'être des arguments plausibles. L'avantage escompté est un gain de temps.

Un autre avantage de l'approche à partir des prédications spécialisées est que le seul investissement à faire est une compétence en langue de spécialité. Or c'est bien ce qu'exigent de toute façon à la fois la pédagogie des langues, celle des disciplines scientifiques et techniques en français, la traduction spécialisée, la lexicographie spécialisée et l'harmonisation terminologique.

S'il faut qualifier cette auto-formation par le haut-débit, rendue possible dans le cas de langues bien documentées sur le Web comme l'est le français, on pourrait parler de terminologie transtextuelle.

BIBLIOGRAPHIE

- Allwood (J.), Andersson (L.-G.) et Dahl (Ö.), 1989 : *Logic in linguistics*, 8^e édition, Cambridge, New-York et Melbourne, Cambridge University Press.
- Aussenac-Gilles (N.), Condamines (A.), 2004 : « Documents électroniques et constitution de ressources terminologiques ou ontologiques », dans *Information – Interaction – Intelligence*, 4-1, pp. 75-94.
- Bessé (B. de), 2000 : « Le domaine », dans Thoiron (Ph.) et Béjoint (H.), édit., *Le Sens en terminologie*, Lyon, Presses universitaires de Lyon (Travaux du C.R.T.T.), pp. 182-197.

- Bogacki (K.), 1992 : « Pendant revisité », dans *Études de linguistique romane et slave*, Cracovie, ENS, pp. 69-78.
- Bourigault (D.), Aussenac-Gilles (N.) et Charlet (J.), 2004 : « Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unifié pour trois études de cas », dans *Revue d'intelligence artificielle*, 18-1, pp. 87-110.
- Cohen (B.), 1986 : *Lexique des cooccurrents. Bourse, conjoncture économique*, Montréal, Linguattech.
- Dancette (J.) et L'Homme (M.-C.), 2005 : « Building specialized dictionaries using lexical functions », dans *Linguistica Antverpiensia*, 3/2004, pp. 113-131.
- Felber (H.) et Budin (G.), 1989 : *Terminologie in Theorie und in Praxis*, Tübingen, Gunter Narr.
- Grabar (N.) et Hamon (T.), 2004 : « Les relations dans les terminologies structurées : de la théorie à la pratique », dans *Revue d'intelligence artificielle*, 18-1, pp. 57-85.
- Gross (G.) et Mathieu-Colas (M.), 2001 : « Description de la langue de la médecine », dans *Meta*, XLVI-1, pp. 68-81.
- Gross (G.) et Guenther (F.), 2002 : « Comment décrire une langue de spécialité ? », dans *Cahiers de lexicologie*, 80, 2002-1, pp. 179-199.
- Heid (U.) et Freibott (G.), 1991 : « Collocations dans une base de données terminologiques et lexicales », dans *Meta*, 36-1, pp. 77-91.
- Karolak (S.), 1989 : *L'article et la valeur du syntagme nominal*, Paris, Presses universitaires de France.
- Kerremans (K.), 2005 : « Categorisation frameworks in Termontography », dans *Linguistica Antverpiensia*, 3/2004, pp. 263-277.
- Kocourek (R.), 1991 : *La langue française de la technique et de la science*, 2^e éd., Wiesbaden, Oskar Brandstetter.
- Lame (G.), 2003 : « Using text analysis techniques to identify legal ontologies components », dans ICAIL 3003 Workshop on Legal Ontologies & Web Based Legal Information Management, 9 p.
- Le Moigno (S.), Charlet (J.), Bourigault (D.) et Jaulent (M.-C.), 2002 : « Construction d'une ontologie à partir de corpus : expérimentation et validation dans le domaine de la réanimation chirurgicale » dans *Actes des 13^{es} journées francophones d'ingénierie des connaissances (IC 2002)*, Rouen, pp. 229-238.
- Lerat (P.), 1988 : « Terminologie et sémantique descriptive », dans *La Banque des mots*, n° spécial CTN, pp. 11-30.
- Lerat (P.), 1995 : *Les Langues spécialisées*, Paris, PUF ; trad. esp. *Las Lenguas especializadas* (1997), Barcelone, Ariel.
- Lerat (P.), 2005 : « Web et terminologie philologique », dans *Linguistica Antverpiensia*, 3/2004, pp. 231-244.
- L'Homme (M.-C.), 2004 : *La Terminologie. Principes et techniques*, Montréal, Presses de l'Université de Montréal.
- L'Homme (M.-C.), Heid (U.) et Sager (J.-C.), 2003 : « Terminology during the past decade (1994-2004) : An Editorial statement », dans *Terminology*, 9 : 2, pp. 151-161.
- Mejri (S.), 1997 : *Le figement lexical*, Tunis, Publications de la faculté des lettres de la Manouba.
- Nowakowska (M.), 1992 : « Interprétation de la manière dans les noms abstraits », dans *Études de linguistique romane et slave*, Cracovie, ENS, pp. 437-448.
- Pavel (S.) et Nolet (D.), 2001 : *Précis de terminologie/Handbook of Terminology*, Ottawa, Travaux publics et services gouvernementaux du Canada.
- Rastier (F.), 2004 : « Ontologie (s) », *Revue d'intelligence artificielle*, 18-1, pp. 15-40.
- Sager (J.-C.), 1990 : *A Practical Course in Terminology Processing*, Amsterdam, John Benjamins.
- Sager (J.-C.), 2000 : « Pour une approche fonctionnelle de la terminologie », dans Thoiron (Ph.) et Béjoint (H.), édit., *Le sens en terminologie*, Lyon, Presses universitaires de Lyon (C.R.T.T.), pp. 40-60.
- Sánchez-Gijón (P.), 2005 : « La extracción de conocimiento y terminología a partir de corpus ad hoc : el uso de documentos digitales de la web pública », *Linguistica Antverpiensia*, 3/2004, pp. 179-202.

- Slodzian (M.), 2000 : « L'émergence d'une terminologie textuelle », dans Thoiron (Ph.) et Béjoint (H.), éd., *Le sens en terminologie*, Lyon, Presses universitaires de Lyon (Travaux du C.R.T.T.), pp. 61-85.
- Temmerman (R.), 1997 : « Questioning the univocity ideal. The difference between sociocognitive Terminology and traditional Terminology », dans *Hermes. Journal of Linguistics*, 18, pp. 51-91.
- Zweigenbaum (P.), 1999 : « Encoder l'information médicale : des terminologies aux systèmes de représentation des connaissances », dans *Innovation stratégique en information de santé*, n° 2-3, pp. 27-47.

LE TRAVAIL TERMINOLOGIQUE POUR LA TRADUCTION DE L'HISTOIRE

Ana Escartín Arilla
Université de Saragosse
Saragosse, Espagne
anaescar@unizar.es

Résumé : *La traduction précise des termes est un aspect-clé dans la traduction de textes scientifiques. La traduction de textes d'histoire, en tant que traduction spécialisée, requiert l'accès à des matériaux terminographiques adéquats aux problèmes qui lui sont propres et qui sont, par la nature même des sciences humaines, principalement extralinguistiques. On envisage l'élaboration de dictionnaires bilingues de termes d'histoire pour les traducteurs, dont le but principal serait précisément l'explicitation des problèmes procédant de la transmission entre cultures, plutôt qu'entre langues.*

Mots-clés : *histoire, terminographie, langue spécialisée, notions historiographique*

1. INTRODUCTION

Les sciences humaines, et plus particulièrement l'histoire, ont été traditionnellement négligées par la terminologie, plutôt consacrée aux champs de la recherche dont la nature scientifique n'a jamais été mise en question, d'où le manque d'ouvrages monographiques de terminologie historique et même de références systématiques à l'histoire dans des textes généraux portant sur des questions terminologiques. D'autre part, les caractéristiques inhérentes à l'histoire comme discipline en ce qui concerne les méthodes de recherche, les présupposés théoriques, les formes du discours et de l'argumentation, ainsi que la nature même de son sujet d'étude, entraînent des difficultés spécifiques en matière terminologique.

Malgré ses particularités, et malgré l'énorme et croissante quantité de livres et articles écrits et traduits dans ce domaine, la traduction du domaine des sciences humaines n'est pas habituellement considérée comme une discipline spécifique dans le monde académique, comme le sont la traduction juridique, économique ou technique et, en général, il y a très peu de traducteurs « spécialisés » et surtout « formés » en histoire, histoire de l'art, anthropologie, etc. En outre, les traducteurs qui doivent faire face à des textes portant sur ces matières ne peuvent pas compter sur des matériaux terminographiques adéquats et précis. Ils ont bien entendu accès à des glossaires, des dictionnaires monolingues spécialisés – aussi très peu nombreux – et des ouvrages monographiques, mais ils manquent de dictionnaires bilingues et plurilingues qui résolvent des difficultés terminologiques concrètes. Il faut noter que, dans une matière telle que l'histoire, la traduction précise des termes

spécialisés est un aspect clé qui se trouve en marge des débats autour de la fidélité, puisqu'il s'agit d'unités lexicales univoques qui tendent à la littéralité référentielle (Cabré 2002) et posent donc ce que l'on connaît comme « problèmes de traduction », c'est-à-dire, des difficultés objectives (linguistiques, extralinguistiques, méthodologiques) apparues lors du processus de traduction (Hurtado 2001 : 286). Un dictionnaire bilingue destiné aux traducteurs devrait viser à résoudre non seulement les problèmes linguistiques mais aussi les extralinguistiques, lesquels tombent les traducteurs de textes spécialisés en histoire ou en historiographie.

Cette communication va donc essayer de systématiser les éléments qui font de l'histoire une matière particulièrement conflictuelle lorsqu'il s'agit d'élaborer des travaux terminographiques et qui doivent être considérés par les équipes de travail visant à créer un hypothétique dictionnaire bilingue ou plurilingue pour la traduction spécialisée. Dans le but de cataloguer ces éléments, on va examiner les différentes phases du travail terminographique pour y repérer les aspects différenciant significativement de ceux qui sont caractéristiques de ce qu'on appelle les « sciences exactes ». Mais nous ferons d'abord une brève réflexion à propos de la question suivante : L'historiographie requiert-elle une traduction « spécialisée » ?

2. LA NÉCESSITÉ D'UNE TRADUCTION « SPÉCIALISÉE » EN MATIÈRE HISTORIQUE

Existe-t-il un « langage de l'histoire » ? L'histoire, en tant que science, possède une terminologie propre et caractéristique, créée – à travers des mécanismes divers – et employée par les historiens pour communiquer leurs connaissances. Nonobstant, étant donné la nature de cette matière, et des sciences humaines en général, dont l'objet d'étude sont les créations humaines à travers le temps, un grand nombre des termes employés sont des mots appartenant à la réalité – c'est-à-dire attribués aux objets et aux événements par une communauté humaine – qui deviennent des concepts théoriques lorsqu'ils sont employés par les scientifiques pour expliquer des phénomènes historiques (Marrou 1985 : 101-102). Comme l'explique Julio Aróstegui, « [il] existe à peine des termes *construits historiographiquement* pour désigner des phénomènes spécifiques », ce qui n'implique pas une sous-estimation de la précision du langage employé par l'histoire. (Aróstegui 1995 : 28-29). On parle alors plutôt d'« usage spécialisé du langage » (Ciapusco et Kuguel 2001) ou de « sens situé » des termes (Cabré 2001), c'est-à-dire que les termes à sens général acquièrent un sens spécialisé en fonction de leur usage et aussi en fonction d'une signification nouvelle – plus pratique et plus éloquente, mais en même temps plus partielle (Marrou 1985 : 102) – assignée par les théoriciens. Les facteurs qui détermineraient la spécialisation d'un terme quelconque seraient, selon Cabré, une thématique spécialisée, des usagers spécialistes et des situations formelles de communication, réglées par des critères professionnels ou scientifiques, des normes auxquelles les termes employés par les historiens répondent sans aucun doute (Cabré 2001). En tout cas, on peut cataloguer le « langage de l'histoire » comme un langage scientifique en tant que « mécanisme employé pour la communication, dont l'univers se situe dans tout champ de la science, cette communication se pro-

duisant soit exclusivement entre spécialistes, soit entre eux et le grand public, quelle que soit la situation communicative et le canal choisi pour l'établir » (Gutiérrez Rodilla 1998 : 20)

De la même façon, Ciapusco et Kuguel (2001) parlent de l'inexistence d'un clivage net entre texte général et texte spécialisé, et affirment la difficulté de déterminer les critères qui justifient que l'on considère un texte comme spécialisé. Plusieurs théoriciens ont essayé de démontrer que le critère thématique n'est pas décisif dans la détermination du degré de spécialisation d'un texte. Il semble plausible que ce soit le critère du « contrôle conceptuel » (Cabré 2002) qui rend possible une telle distinction. Cependant, il paraît aussi évident que, dans le cas de l'élaboration d'un dictionnaire comme celui qui constitue notre hypothèse de travail, le critère thématique est décisif, tandis que le niveau de « contrôle conceptuel » du texte est une question assez difficile à cerner, bien qu'elle ne puisse pas être négligée lors de la sélection des matériaux qui constitueront le *corpus* du travail terminographique.

Malgré ces difficultés de délimitation de la notion de « texte spécialisé », dans le domaine de la traduction il existe depuis presque toujours une distinction nette entre « traduction générale » et « traduction spécialisée » qui est toujours présente dans le travail du spécialiste, dès le moment où les tarifs appliqués varient selon le degré de spécialisation du texte original. En tout cas, la traduction de textes d'histoire peut être considérée comme une traduction spécialisée – sans oublier bien entendu les différents niveaux de difficulté des textes – étant donné précisément les caractéristiques particulières de sa terminologie, que l'on analysera attentivement plus tard.

3. LE TRAVAIL TERMINOGRAPHIQUE : DIFFICULTÉS DE L'ÉLABORATION D'UN DICTIONNAIRE BILINGUE D'HISTOIRE

3.1. Les matériaux du travail préalable

Tandis que dans les matières techniques il existe toute une tradition de dictionnaires, lexiques et vocabulaires auxquels s'adresser comme soutien de travail et qui ont été réalisés du point de vue de la terminologie, ce sont en général les historiens qui ont dû faire face au manque de ce genre d'ouvrages de référence, en publiant des catalogues de termes historiques qui, normalement, ne respectent pas les règles du travail terminologique et qui ont des buts autres que l'appui au travail du traducteur. D'habitude, les dictionnaires de termes ou de concepts historiques, décrivent les notions avec les traits caractéristiques au discours historique, d'autant plus que la matière traitée y prête spécialement. En plus, ils répondent parfois à la volonté de l'auteur de défendre à travers la définition des termes une vision particulière des processus historiques, ayant donc une vocation nettement prescriptive. Les implications du choix des termes sont, comme on verra plus tard et sauf dans des cas comme le côté plus technique de l'histoire de l'art, une des questions les plus conflictuelles en cette matière et la réflexion à propos de cela fait l'objet de nombreux ouvrages historiographiques.

3.2. La délimitation du thème

Il semble évident qu'un hypothétique « dictionnaire d'histoire français-espagnol » ne serait pas suffisamment exhaustif ou suffisamment pratique, même pas un dictionnaire de l'histoire d'un pays concret. En règle générale, les ouvrages monolingues dans ce domaine portent soit sur un moment ou un phénomène historique précis, soit sur un aspect concret de la société humaine analysé tout au long du temps, éléments qui déterminent le thème du travail terminographique. Ainsi trouve-t-on des études sur « la transition espagnole », « la sorcellerie » ou « les révolutions bourgeoises ». Dans le cas d'un éventuel dictionnaire bilingue, l'ampleur du sujet, ainsi que le grand nombre de termes associés à celui-ci, obligerait à une limitation temporelle ou thématique très précise, ou bien à l'élaboration d'une encyclopédie complète.

3.3. Le choix du *corpus*

Quand on parle de l'histoire, on affronte un problème de définition, étant donné la confusion terminologique entre la science et son objet. Julio Aróstegui a proposé l'expression d'« amphibologie » du terme « histoire » pour faire référence au fait qu'« un seul mot, *histoire*, a traditionnellement désigné deux choses différentes : l'histoire comme réalité dans laquelle l'homme est inséré et, d'autre côté, la connaissance et l'enregistrement des situations et des événements qui signalent et manifestent cette insertion. » (Aróstegui 1995 : 20). Le terme « Historiographie » est à son tour employé à certains moments comme synonyme d'histoire, et à d'autres pour désigner l'ensemble des ouvrages rédigés par les historiens à propos de leur objet d'étude (Aróstegui 1995 : 23-27). On peut donc faire une différenciation entre les documents historiques, ceux qui ont été rédigés tout au long de l'histoire par les hommes avec des buts tout à fait variés et qui font l'objet, sans en avoir vocation, d'une étude historiographique *a posteriori*, et les textes d'historiographie, c'est-à-dire, d'une part, ceux qui sont le résultat d'une analyse des faits passés, créés par des spécialistes avec des objectifs tout d'abord scientifiques, et d'autre part, ceux dont la matière est l'étude même de la recherche historiographique. Pour éviter la confusion, on peut parler de « documents historiques » dans le cas des sources primaires employées par les historiens dans leur recherche d'une part, et de « textes spécialisés en histoire » et/ou « textes spécialisés en historiographie » d'autre part, qui sont en principe ceux qui doivent former le *corpus* du travail terminographique qui nous occupe.

Or, la distinction entre document historique et texte spécialisé est quand même diffuse dans certains cas. La conception de l'histoire comme discipline a beaucoup changé, et les textes écrits il y a longtemps par des historiens sont vus actuellement comme plus proches de la littérature que de la science et très souvent analysés comme des sources historiques en elles-mêmes. C'est le cas des auteurs classiques et on le voit aussi très clairement à propos des débats, entre autres, autour de la Révolution française. L'introduction à *La Révolution française* d'Albert Soboul (1982) est un exemple d'exploitation d'ouvrages d'historiens français contemporains constituant des sources pour l'analyse.

D'ailleurs, la notion de « texte spécialisé en histoire » n'est pas tellement indubitable. Il y a certes des écrits sortis de la plume d'un historien – ou en tout cas d'un spécialiste – et pourvus d'éléments caractéristiques du discours historiographique. Cependant, on peut trouver d'autres textes qui, par des raisons diverses, peuvent être considérés aussi comme historiographiques. C'est le cas de certains textes appartenant à d'autres disciplines scientifiques qui apportent des données historiques ou même adoptent les éléments propres à la narration historiographique. En ce sens, il existe depuis longtemps un processus d'interaction entre les sciences (sociologie, politologie, droit, économie) qui a transformé de manière radicale la structure narrative de leurs discours respectifs et auquel l'histoire n'a pas été étrangère. Carreras Ares (2000 : 215-257) a effectué une révision exhaustive de ce processus dans le cas de l'histoire.

Une des conditions qui doivent être remplies par les documents constituant le *corpus* est leur actualité, étant donné la nature changeante des connaissances scientifiques. Dans le cas de l'histoire, le moment – et le lieu – où un certain texte a été produit détermine non seulement l'état des connaissances mais aussi la façon d'envisager le sujet. Si une telle affirmation peut être appliquée à l'ensemble des sciences, personne ne niera que la matière historique est particulièrement sensible à des questions de nature idéologique au-delà de l'observation rigoureuse des faits. La terminologie employée dans les textes d'histoire n'est pas neutre et la sélection des documents qui serviront de *corpus* à notre travail de recherche de termes spécialisés doit tenir compte des facteurs extralinguistiques déterminant le choix d'une terminologie concrète.

D'autre part, les documents doivent être originels, c'est-à-dire, rédigés dans la langue du travail. L'histoire, comme la plupart des sciences humaines, axe très souvent ses analyses sur un pays concret qui présente des caractéristiques particulières ayant reçu une désignation précise dans la langue du pays concerné. Ainsi, selon le thème déterminé auparavant, il faudra tenir compte des difficultés terminologiques que les références culturelles – propres aux sciences humaines – vont poser. En outre, parfois le chercheur étudie le cas d'un contexte géographique dont la langue n'est pas la sienne. Un hispaniste, par exemple, doit entamer son propre processus de traduction des références culturellement spécifiques de l'Espagne pour les transmettre à son public, qui parle une langue autre que l'espagnol. Le terme « cacique », en espagnol, poserait de graves problèmes de transmission, par exemple, à un historien anglais, qui devrait trouver un mot qui transmette le sens spécifique du terme dans la réalité espagnole, en plus d'expliquer avec précision le contenu significatif de celui-ci, un problème, par ailleurs, que doit aussi affronter l'historien espagnol, puisqu'il s'agit d'une issue conceptuelle. Dans ce but, l'hispaniste devra employer des stratégies diverses que l'on analyse ailleurs et dont le traducteur vers d'autres langues – y compris dans ce cas l'espagnol – devra prendre en considération lorsqu'il s'adresse à un nouveau public. On va présenter un cas très clair portant sur des concepts de l'historiographie. Carreras Ares (2000 : 134-142) étudie le phénomène de la régionalisation de l'historiographie à partir des années soixante du XX^e siècle, et

parle de trois réalités : « histoire régionale », « *Landesgeschichte* » et « *Historia regional* ». Il s'agit, comme on voit, d'un phénomène qui a eu lieu plus ou moins dans la même époque dans des pays différents, mais le terme employé dans chaque pays a des connotations particulières. Ce n'est pas seulement une question de langue : on ne peut pas traduire « histoire régionale » par « *Landesgeschichte* » parce que ses mots font référence à des notions différentes, bien que parallèles et comparables. Le *corpus* définitif de notre éventuel travail terminographique devrait être représentatif de toutes les réalités exposées.

3.4. Le dépouillement des documents

Dans la première phase du travail terminographique avec les documents, il faut faire face à l'une des plus grandes difficultés : la sélection précise des mots qui peuvent être considérés comme des termes de la matière historique. Il y a des termes employés depuis longtemps par les historiens et qui décrivent des phénomènes nettement historiques, tels que « néolithique », « féodalisme » ou « guerre froide », des termes qui peuvent être employés dans d'autres champs, comme l'économie, la sociologie ou le journalisme, mais qui appartiennent sans aucun doute à la tradition historiographique. Mais les sciences humaines comprennent l'ensemble des aspects de la vie humaine, d'où la nature essentiellement interdisciplinaire de leurs termes. On pourrait se demander si le terme « souveraineté nationale » est proprement historique. Procédant du champ juridique ou même de la politologie, ce terme est devenu, par l'effet des événements historiques, un terme fondamental pour la compréhension de certains processus du point de vue de l'histoire. L'adéquation de l'attribution d'un terme concret à une matière scientifique dépend en tout cas du contexte d'usage.

Le fondement des « langues de ou en spécialité » est certes l'emploi de termes de la langue générale dans la communication spécialisée (Ciapusco et Kugel 2001, Cabré 2002), et de ce point de vue rien ne distingue l'histoire des sciences exactes. Or, dans les sciences humaines, on trouve rarement des « unités terminologiques » des types décrits dans les manuels de terminographie, celles qui peuvent en principe être reconnues par leurs composants formels, tandis que des termes tels que « propriété de la terre » ou « État », difficiles à repérer à travers des systèmes automatiques de traitement de l'information, ne devraient jamais être négligés. Le contexte joue à nouveau un rôle constitutif.

Il existe encore un troisième aspect : celui du « vocabulaire d'appui », des mots et des expressions, selon Darbelnet, caractéristiques des textes spécialisés, mais ne répondant pas aux normes de la terminologie (Lerat 1995 : 52). Par exemple, l'expression « propriété de la terre » peut être considérée comme un « terme » dans le contexte de la science historique ou plutôt comme un mot très employé mais non suffisamment spécialisé. Dans la décision de la faire entrer dans notre éventuel fichier terminologique, le facteur subjectif joue un rôle indéniable, mais il est quand même, précisément pour cette raison, indispensable d'établir des critères préalables qui dirigent la sélection des termes significatifs.

Mais le problème ne se pose pas tellement lors du dépouillement, c'est-à-dire, du repérage des mots significatifs pour le thème proposé, mais plutôt lors de l'attribution d'une définition à un terme ou, à l'inverse, l'assignation d'un terme précis à une notion, c'est-à-dire, lors de la préparation du fichier terminologique.

3.5. L'élaboration du fichier terminologique

La question la plus problématique dans ce point est, comme on vient de dire, la définition de la notion référée par le terme. Les notions dans la science historique, ainsi que dans les autres sciences humaines, ne sont en général pas aussi « tangibles » que dans les sciences « exactes » et surtout les techniques. Celles qu'on a déjà mentionnées, présentes dans tout essai de catalogage du vocabulaire historique, sont suffisamment abstraites et ont assez d'implications pour générer des conflits parmi les spécialistes à propos de leur contenu, et même le choix du terme concret pose des difficultés et génère des débats. Des facteurs non strictement scientifiques interviennent à leur tour dans ce type de délibérations. Le cas de la controverse autour de la nature du régime franquiste, résumée par Sevillano Calero (1999), est un exemple clair de débat terminologique entraînant des questions idéologiques.

On trouve aussi de graves difficultés lorsqu'il faut trouver une équivalence des termes dans d'autres langues, dont on a déjà parlé quand on a abordé les conditions des documents du *corpus*. Chaque langue a déjà accordé, avec plus ou moins de succès, une unité terminologique précise aux notions qui appartiennent à l'histoire d'une communauté considérablement large, telles que « révolution industrielle » ou « paléolithique ». Or, même dans le cas de ces notions, parfois le terme employé pour la première fois est si éloquent – c'est le cas de « *putting-out system* » en anglais – que les historiens décident de ne pas le traduire dans d'autres langues, dans le but de conserver son sens propre. Dans d'autres occasions, des expressions déjà consolidées font l'objet d'un nouveau débat à un moment donné, dû aux progrès dans les recherches ou à l'apparition de visions d'ensemble différentes tout au long du temps. C'est le cas de la révision de la notion de « *Reconquista* » dans l'Espagne du Moyen Âge (Barbero et Vigil 1979). D'un autre côté, lorsqu'il s'agit de notions culturellement ou géographiquement restreintes, leur traduction peut entraîner des confusions ou des imprécisions théoriques. La dénomination assignée aux réalités devient dans ce cas-là une question épineuse et les options choisies varient selon l'auteur ou selon le cas concret. Elles oscillent entre la conservation du terme dans la langue originelle (les « sans-culottes » français), la création d'un terme dans la langue du spécialiste qui ressemble à l'original (le terme « caciquisme » pour parler du « *caciquismo* » espagnol) ou la sélection d'un mot de la langue du chercheur et l'attribution à ce mot d'un nouveau sens culturellement déterminé (« *day labourer* » pour « *jornalero* »). Voilà une des plus grandes difficultés de la traduction de l'histoire. Et la question se complique si l'on doit faire face à un terme qui, comme on a déjà signalé, a souffert à son tour d'un processus de traduction de la part de l'historien, qui a essayé de trouver un mot « équivalent » dans sa propre langue pour reproduire le sens, par exemple, de « *Fuero* », en espagnol, en tant que compilation particulière de lois. Il est alors

indispensable de repérer la désignation dans la langue originale pour répondre, comme traducteurs, à la réalité historique concernée. Ainsi convient-il de faire attention à la notion d'« équivalence dynamique », proposée à l'origine par Nida et très bien accueillie par les spécialistes, qui entraîne la priorité de la conformité contextuelle et de l'adéquation aux besoins des récepteurs selon le genre du texte (Hurtado 2001) dans le choix du terme définitif.

4. CONCLUSION

La traduction des textes d'histoire est une traduction spécialisée qui, en tant que telle, requiert, d'une part, des professionnels maîtrisant les notions historiographiques fondamentales et les particularités de la science historique, ainsi que les stratégies et difficultés de l'activité traductrice, et d'autre part, des matériaux de travail consacrés à résoudre les problèmes spécifiques qu'elle pose. La systématisation de tels problèmes n'est pas complète, mais l'énumération effectuée ici nous rapproche de la réalité à laquelle doivent faire face les traducteurs spécialisés en sciences humaines et ouvre une voie de travail pour les terminologues dans le but de répondre convenablement à leurs nécessités. Les dictionnaires bilingues généraux ne peuvent aider le traducteur à affronter les difficultés dont on a essayé d'élaborer un catalogue. Voilà pourquoi s'impose la nécessité d'élaborer des dictionnaires spécialisés en matière historique, développant des explications précises à propos des difficultés linguistiques et surtout extralinguistiques du choix des termes dans les textes spécialisés en histoire et historiographie. La nouveauté et la puissance de l'ouvrage que l'on envisage résident précisément dans la capacité à résoudre des problèmes extralinguistiques.

BIBLIOGRAPHIE :

- Aróstegui (J.), 1995 : *La Investigación histórica : teoría y método*, Barcelona, Crítica.
- Barbero (A.), Vigil (M.), 1979 : *Sobre los orígenes sociales de la Reconquista*, Barcelona, Ariel.
- Cabré (M.-T.), 1993 : *La Terminología. Teoría, Metodología, Aplicaciones*, Barcelona, Editorial Antártica/Empúries.
- Cabré (M.-T.), 2002 : « Textos especializados y unidades de conocimiento : metodología y tipologización », dans García Palacios (J.), Fuentes Morán (M. T.), édit. : *Texto, Terminología y Traducción*, Salamanca, Ediciones Almar, pp. 15-36.
- Carreras Ares (J.-J.), 2000 : *Razón de historia. Estudios de historiografía*, Madrid, Marcial Pons.
- Ciapusco (G.), Kuguel (I.), 2001 : « El discurso especializado : propuesta de tipologización », dans *Actas del II coloquio de investigadores en estudios del discurso*, La Plata.
- García Palacios (J.), Fuentes Morán (M.-T.), édit., 2002 : *Texto, Terminología y Traducción*, Salamanca, Ediciones Almar.
- Gutiérrez Rodilla (B.-M.), 1998 : *La Ciencia empieza en la palabra. Análisis e Historia del lenguaje científico*, Barcelona, Península.
- Hurtado Albir (A.), 2001 : *Traducción y Traductología : Introducción a la traductología*, Madrid, Cátedra.
- Lerat (P.), 1995 : *Les Langues spécialisées*, Paris, PUF.
- Marrou (H.-I.), 1985 : *Del conocimiento histórico*, Buenos Aires, Per Abbat Editora.

- Sager (J.-C.), 1993 : *Curso práctico sobre el procesamiento de la terminología*, Madrid, Pirámide.
- Sevillano Calero (F.), 1999 : « Totalitarismo, fascismo y franquismo : el pasado y el fin de las certidumbres después del comunismo », dans Moreno Fonseret (R.), Sevillano Calero (F.) : *El Franquismo. Visiones y balances*, Alicante, Publicaciones de la Universidad de Alicante, pp. 14-26.
- Soboul (A.), 1982 : *La Révolution Française*, Paris, Arthaud.

ENTRE PROTOTYPISATION ET MISE EN DISCOURS : LES ENJEUX DU SENS

Abdou Elimam
ENSET-Oran
Oran, Algérie
LESCLaP (U. d'Amiens)

Résumé : *Le concept de prototypisation dans son héritage lexicaliste, se voit interrogé à partir du concept de « praxème » (R. Lafont). De même que le « contexte discursif » se voit mettre en contraste avec le « contexte situationnel, voire pragmatique ». Les procédures d'analyse automatique et de calcul du sens par l'humain seront également distinguées pour restituer la dynamique du sens au parcours discursif, au texte.*

Mots-clés : *prototype, praxème, signification, parcours discursif, statut discursif, cohésion discursive*

Dans un ouvrage récent, édité par Nicole Delbecque (2002), se trouve défendu le concept de *prototypisation*. Nous nous proposons de vérifier la pertinence d'un tel concept¹ en lui opposant la démarche praxématique (Lafont 1978). La question du contexte intervient, selon les cas de figure, de manière plus ou moins explicite. En effet, le contexte discursif peut-il générer les mêmes retombées significationnelles que le contexte situationnel, voire pragmatique ? La tendance de ces deux dernières décennies en linguistique témoigne de rapprochements substantiels entre différentes écoles de pensée. Précisément autour du poids qu'instaure le contexte dans le calcul de la valeur significationnelle, produite en discours. On sait que ces questionnements ont pris de l'ampleur dans les démarches d'analyse automatique du discours puis dans celles du traitement automatique des langues naturelles. Elles font écho en « linguistique du *corpus* », en « grammaire du texte » et, plus généralement, en linguistique des opérations énonciatives et prédicatives. Dès ses premières formulations majeures, la linguistique praxématique entamait un tel tournant épistémologique. Le courant culiolien², de son côté, œuvrait dans le même sens. Et bien d'autres courants linguistiques (que l'on songe aux travaux de M. A. K. Halliday, par exemple) se sont risqués dans cette voie de la détermination du contexte dans le calcul significationnel. Y compris le courant générativiste contemporain, qui s'est vu contraint d'intégrer de telles contraintes³. L'objectif que nous nous assignons, dans le présent travail, consiste à appréhender la question du sens à par-

¹ Concept dont le mode opératoire s'apparente incontestablement à celui de l'approche lexicaliste.

² Cf. Culioli en bibliographie.

³ Que l'on songe à ces « traces de mouvement » et aux contraintes imposées par les « rôles thématiques », par exemple. Cf. Pollock (1997), en bibliographie.

tir de deux angles : l'angle « lexicaliste » d'une part, et celui de l'approche discursive de l'autre. Qu'il nous soit permis, le temps de poser notre problématique, d'identifier :

1. L'approche lexicaliste à une vision modulaire qui range le lexique dans une sorte de dictionnaire interne. Le traitement du dictionnaire est séquentiel et les instances (ou entrées) sont saisies à partir de traits à la fois sémantiques, phonologiques et morphosyntaxiques. Le dictionnaire peut, si nécessaire, intégrer des traits pragmatiques.

2. Quant à l'approche discursive, elle sera appréhendée essentiellement comme une dynamique mettant en œuvre des opérations linguistiques destinées à un sujet interprétant pour en calculer la signification à partir d'un sens préposé.

En somme, dans un cas le sens est un « déjà là » que l'agencement syntaxique confirme, dans l'autre, le sens est l'aboutissement d'un calcul que l'agencement syntaxique commande. Cependant, dans les deux cas, l'agencement syntaxique joue un rôle certain. Il arrive souvent que ces deux approches soient réduites à une opposition entre une linguistique de la langue, d'une part, et une linguistique du discours (ou de la parole), d'autre part. Dans le corps épistémologique où s'inscrivent les tendances contemporaines en linguistique cognitive, la démarche appréhendant le sens prend précisément appui sur la valeur prototypale du sens des « mots ».

À titre d'illustration⁴, on attribue à une catégorie lexicale comme « chaise » un tas de sous-types (p. ex. *chaise de cuisine/d'enfant/de bureau, berceuse, fauteuil, fauteuil roulant*). Certains d'entre eux sont considérés comme étant de « meilleurs représentants de la classe que d'autres ». C'est précisément le membre le plus représentatif, « le meilleur » de la catégorie, qui est appelé « le *prototype*, ou membre *prototypique* de la catégorie... Celui qui nous vient spontanément à l'esprit »... Les autres membres de la catégorie, ceux qui sont moins courants voire moins représentatifs, sont dits *marginiaux* ou *périphériques*. Les référents prototypiques présentent donc une plus grande saillance. Notons cependant que, « en général, le noyau ou le centre d'une catégorie lexicale est bien fixé et clairement défini. Par contre, les limites en restent plutôt vagues et elles ont tendance à empiéter sur celles d'autres catégories lexicales ». Ainsi en est-il par exemple des catégories apparentées telles que « chaise » et « tabouret ».

L'approche est certes fidèle à des constats statistiquement motivés. Cependant, ce qui la distingue malgré tout, c'est la vision du fonctionnement des langues qu'elle dégage. Les « catégories lexicales » sont posées comme un des éléments, voire modules, constitutifs de cet ensemble appelé « langue ». Faire sens consiste donc à mettre en relation – *via* le module « syntaxe » – des sélections lexicales basées sur un type de pointage : prototypique, marginal, ou bien périphérique. Ce sont donc des « déjà là » que nous sollicitons en vue de traduire – *via* l'activité langagière – une représentation donnée.

⁴ Nicole Delbecq (2002) : *op. cit.* pp. 35-36.

Pour nous qui avons décidé de débattre du rapport qu'entretient le « sens » au « contexte », la question centrale à nos yeux, est précisément celle de savoir si le sens est un donné étanche que le contexte confirme, atténue ou renforce. Ou bien, si le sens est potentialité significationnelle dont le réglage demeure social (voire dialogique), c'est-à-dire réservé à l'interprétation. Quant au contexte, il est soit discursif, soit situationnel (ou pragmatique). Et l'impact de l'un ou de l'autre sur l'interprétation n'est pas identique. En effet, le contexte d'un conte de fées peut permettre des réglages de sens bien plus libres que celui d'un procès-verbal instruit par un fonctionnaire. Par contre le contexte discursif renvoie à la notion de cohérence et de cohésion textuelles qui fonctionnent comme un maillage interne contraignant la forme externe de la production langagière ; sa morpho-phonologie, en somme.

Il se trouve que la linguistique praxématique, telle qu'inaugurée par Robert Lafont, apporte bien des réponses innovantes et heuristiques à nos interrogations. En effet la démarche praxématique forge ses formulations autonomes du retournement du signe saussurien en une praxis signifiante. Au lieu du signifié, le praxème est puissance à signifier, plutôt. Mais également en reconnaissant au sujet désirant et parlant la place centrale dans les procès opératoires commandant toute production langagière. Ces opérations linguistiques (et métalinguistiques) qui avaient déjà été restituées dans leur dynamique interne par Gustave Guillaume, sont reprises par la praxématique et appréhendées à partir du temps opératif – temps d'activité nerveuse supérieur. En cela la praxématique fait écho à bien des formulations – certes singulières dans leur forme – de l'approche énonciative et prédicative⁵. Notre propre approche ambitionne d'offrir une synthèse de ces deux dernières théories⁶.

Voyons comment la question du sens peut être cernée de ce point de vue-là. Dans toute communauté linguistique, force est de reconnaître une certaine autonomie de cet espace public qu'est la langue commune. Cependant grand est le risque de succomber à la tentation de considérer cette autonomie comme un ailleurs doté de sa propre énergie... à l'écart des parlants, des producteurs effectifs de toute langue humaine. Dans toute production langagière, « Il y a toujours travail linguistique sous la double contrainte pratique d'une adaptation au réel et d'une expression du moi profond » (*op. cit.*, p. 77), rappelle R. Lafont. C'est en cela que la notion de « signifié » faisant prendre le risque d'induire une sorte de transcendantalisme du sens, il a semblé préférable de poser le praxème à partir d'une dynamique discursive de production du sens. Dynamique qui avait déjà trouvé quelques formulations heuristiques sous l'espèce du « sens puissantiel » chez les tenants de la linguistique psychomécanique, mais également sous celle de « l'actualisation », chez C. Bally ; voire la « signifiante », chez E. Benvéniste.

Pour Lafont, le praxème, cet élément de la praxis humaine inscrit en logosphère, est un instrument dont l'efficacité s'inscrit en « duplication » ; les informaticiens y

⁵ Outre A. Culioli, mentionnons, entre autres, Henri Adamczewski (1991). Notons, également, toutes les similitudes à découvrir chez les partisans de la linguistique « psycho-mécanique ».

⁶ On pourra se reporter à A. Elimam (2003), chapitre V, pour une présentation succincte.

auraient préféré le concept de *récurtivité*. Cette duplication s'effectue à la lisière de la tautologie (*l'avion vole, le laser tranche*) en cela qu'elle restitue la praxis matérielle qui est venue s'y inscrire. Dans un tel cadre, l'accès aux figures rhétoriques procède d'une émancipation de la praxis linguistique par rapport à la praxis matérielle. C'est de la dialectique qui unit la praxis matérielle à la praxis linguistique que vont se dégager en culture, à la fois une faculté de typisation – *végétal* – et une faculté d'analyse aboutissant à une « hiérarchie signifiante » – *feuille, tronc ; voire bosquet, forêt*. Ces découpes sont considérées comme des « programmes de sens ». Ce qui nous amène à appréhender les parcours discursifs en termes de connexions programmatiques faisant intervenir des praxèmes dans des relations arthrologiques représentées par des opérations métalinguistiques. Ces dernières apparaissent en surface sous l'espèce d'unités phonologiques dont les représentations s'allègent du poids de la praxis matérielle au profit de la praxis linguistique. C'est par l'action de ces mots-outils ou métapraxèmes que les jeux du sens peuvent opérer, que les connexions entre découpes sont envisageables. Il s'agit alors de mouvements opérés en termes d'emboîtements et/ou déboîtements praxémiques. Disons, pour simplifier, que les représentations initiales portées par un praxème donné font alors l'objet de « négociations » visant à confirmer ou infirmer l'unité de typisation ou de hiérarchie signifiante. Et c'est du résultat de cette découpe – discursive – que les autres connexions programmatiques sont envisageables. De la sorte, la cohérence est assurée et préservée.

Nous nous proposons donc d'appréhender la question du sens dans sa dynamique pragmatique et discursive, plutôt qu'en termes d'instantiation d'une valeur à partir d'une sélection dans un dictionnaire. Autrement dit, nous envisageons de montrer que le processus de la production du sens, dans le langage humain, n'a recours à un dictionnaire – fût-il mental – qu'allusivement⁷. Nous avons affaire à des espaces discursifs ; des découpes discursives du réel. Et ce sont ces découpes qui vont faire l'objet de parcours où l'on passe toujours du connu vers le moins connu. Du rhématique vers le thématique⁸. Ces mouvements de bascule opèrent par emboîtements – déboîtements de ces mêmes découpes du réel, *i.e.* des virtualités significationnelles dont l'extension peut, en « sortie », être portée et matérialisée par autant de mots et/ou d'expressions et/ou de paradigmes que possible. Le dictionnaire est dans le meilleur des cas une configuration provisoire d'un état de matérialisation d'un potentiel significationnel. Dans le langage en acte, le mot n'est que prétexte à un bain discursif. Un « ticket d'entrée » dont la fonction essentielle est de permettre l'accès à des découpes logosphériques. Or accéder à une découpe logosphérique (p. ex. annonce des titres du journal télévisuel : « L'Irak », « Visite du Pape », « Météo : du vent et de la pluie », etc.) peut tout au plus introduire un espace de cohérence à partir duquel des parcours discursifs, des computations linguistiques, vont pouvoir opérer. Ce sont ces computations qui organisent et construisent le sens visé en sortie.

⁷ La question des « langues spécialisées », comme les appelle P. Lerat (1995), aurait nécessité quelques développements particuliers, qui ne remettent pas en cause la démarche ici préconisée.

⁸ Pour reprendre des concepts clés chez H. Adamczewski, *op. cit.*

Pour rendre compte de notre approche, nous reprendrons un texte illustre et le soumettrons à nos analyses. Il s'agit de l'acte III scène II de *Jules César* de William Shakespeare (1564-1616). Cette scène se déroule après que Brutus et Cassius, tous deux proches de l'Empereur, ont décidé de faire justice en tuant César accusé de constituer, dorénavant, une menace pour la liberté. Antoine, l'un des conspirateurs, cependant non impliqué dans le meurtre, est appelé à dire l'oraison funèbre du défunt. La stratégie oratoire que Shakespeare déploie sous le verbe d'Antoine souscrit à une procédure rhétorique dans laquelle alternent *monologue*, *discours rapporté*, *assertion* et *interrogation* ou *commentaire*. Cette succession offre un avantage certain à l'orateur : il peut s'avancer sur un terrain peu certain, face à un public acquis à la cause des conspirateurs, sans donner l'impression de prendre une position personnelle. Sans donner l'impression de s'impliquer en faveur des uns ou des autres. Il accumule faits et rumeurs sans les opposer de manière frontale ; jusqu'au moment où il sent que son auditoire est prêt à réviser son jugement. Ce retournement de situation constitue un modèle de précaution oratoire, une pièce d'anthologie dans l'écriture littéraire.

Jules César⁹

ACTE III SCENE II — Le Forum.

[...]

ANTOINE

1. Amis, Romains, compatriotes, prêtez-moi l'oreille.
2. Je viens *pour* ensevelir César, *non pour* le louer.
3. Le Mal que font les hommes vit après eux ;
4. Le Bien est souvent enterré avec leurs os ;
5. Qu'il en soit ainsi de César.
6. *Le noble Brutus* vous a dit que César *était ambitieux* :
7. *Si cela était*, c'était un tort grave, et César *l'a gravement expié*.
8. Ici, avec la permission de *Brutus* et des autres –
9. *Car Brutus est un homme honorable* –
10. *Et ils sont tous des hommes honorables* –
11. Je suis venu *pour* parler aux funérailles de César.
12. *Il était mon ami fidèle et juste* ;
13. *Mais Brutus* dit qu'il *était ambitieux* ;
14. *Et Brutus est un homme honorable*.
15. *Il a ramené à Rome nombre de captifs*
16. Dont les rançons *ont rempli* les coffres publics :
17. *Est-ce là ce qui a paru ambitieux dans César ?*
18. *Quand le pauvre a gémi*, César *a pleuré* :

⁹ <http://www.shakespeare-online.com/>

19. L'ambition *devrait être* de plus rude étoffe :
20. *Pourtant Brutus* dit qu'il *était* ambitieux ;
21. *Et Brutus est* un homme honorable.
22. Vous avez tous vu qu'aux Lupercales
23. Je lui *ai* trois fois *présenté* une couronne royale,
24. Qu'il *a* refusé trois fois : *était-ce* là de l'ambition ?
25. *Pourtant Brutus* dit qu'il *était* ambitieux ;
26. *Et assurément, c'est* un homme honorable.
27. Je ne parle *pas pour* contester ce *qu'a* déclaré Brutus,
28. *Mais* je suis ici *pour* dire ce que je sais.
29. Vous *l'avez* tous aimé naguère, *et non sans* motif ;
30. *Quel* motif vous empêche *donc de* le pleurer ?
31. Oh jugement tu *as* fui chez les bêtes *brutes*,
32. Et les hommes *ont* perdu leur raison. Excusez-moi ;
33. Mon cœur *est* dans le cercueil, là *avec* César,
34. Et je dois m'interrompre jusqu'à ce qu'il me *soit* revenu.
35. [...]
36. *Hier encore*, la parole de César *aurait pu*
37. Prévaloir *contre* l'univers : *maintenant* le voilà gisant,
38. *Et il n'est pas* un misérable qui daigne lui faire honneur !
39. *Oh mes maîtres, si j'étais disposé* à exciter
40. Vos cœurs et vos esprits à la révolte et à la fureur,
41. Je *ferais* tort à Brutus et tort à Cassius,
42. Qui, vous le savez tous, *sont* des hommes honorables.
43. Je *ne veux* pas leur faire tort ; *j'aime mieux*
44. Faire tort *au* mort, faire tort à vous-mêmes et à moi,
45. *Que de* faire tort à des hommes si honorables.
46. *Mais voici un parchemin avec le* sceau de César ;
47. Je *l'ai* trouvé dans son cabinet ; *ce sont* ses volontés dernières :
48. [...]

L'objet de cette sublime oraison funèbre consiste en une mise à l'épreuve de l'argument de Brutus contenu dans l'attribut « ambitieux » accolé au nom de César. De la sorte l'attribut d'ambition, supposé trait constitutif du nom César, finit par se dérober sous l'énumération de traits antonymiques, traits s'imposant comme plus représentatifs. En somme nous avons, en entrée, une valeur prototypique – condensée sous l'attribut d'*ambition*. Cette valeur subit une réelle refonte jusqu'à devenir excentrique, voire périphérique. Un tel déplacement significationnel est, en l'espèce, le résultat d'un processus discursif. Comment opère le discours pour amener le lecteur à passer d'une valeur prototypique à une valeur périphérique, voire allusive ?

Commençons par restituer les parcours discursifs de « parler » et d'« ambition » :

1. LA PRISE DE PAROLE (D'ANTOINE) : CE QUE « PARLER » VEUT DIRE

Cette dernière peut se réduire à la trame suivante : « je parle non pas pour le louer, mais pour parler. Non pas pour contester ce que Brutus a déclaré, mais pour dire ce que je sais avant d'arrêter de parler. Faute de quoi, cela ferait du tort à Brutus. »

2. TRAITS SAILLANTS IDENTIFIANT CÉSAR : CE QU'« AMBITION » VEUT DIRE

César est présenté par Antoine comme un « homme parmi les autres hommes ; un ami fidèle et juste ; qui utilise l'apport des rançons pour remplir les caisses publiques ; qui pleure lorsque le pauvre gémit ; qui, par trois fois, a refusé la couronne royale ; qui a été digne de votre amour ; celui dont la seule parole suffisait à défier l'univers. »

3. CE QUE BRUTUS DIT DE CÉSAR : VERS UN VIDAGE DE LA NOTION D'« AMBITION »

Brutus n'est présenté qu'à travers ses déclarations : il dit que « César était ambitieux », il dit son accord pour qu'Antony parle, il dit et redit que « César était ambitieux ».

Certes, Shakespeare met en scène la confrontation d'arguments, présentés comme neutres mais vrais, à des assertions posées comme immuables. Exemple :

« Il était un ami fidèle et juste » vs « Brutus dit que César était ambitieux »

C'est ainsi que tous les arguments apportés par Antoine, mettant au jour des traits de la personnalité de César, sont systématiquement suivis de l'immuable assertion attribuée à Brutus : « César était ambitieux ». La simple mise en rapport des « lexèmes » serait, convenons-en, stérile, à cet égard. Ce qui fait que ces mêmes lexèmes « bougent de l'intérieur », c'est le travail effectué sur eux par des occurrences de nature métapraxémiques. Face aux arguments avancés par Antoine en faveur de César, les assertions attribuées à Brutus sont introduites par *mais* et *pour-tant*. Des connecteurs argumentatifs qui établissent une hiérarchie entre propositions relevant d'un même domaine (cf. cohésion discursive). Les effets de sens du type « concession », « adversativité », « altérité », etc. dérivent de cette hiérarchie interpropositionnelle. Les deux propositions en relation ne peuvent relever de domaines différents, au prix d'un étonnement, d'un éclat de rire, voire d'une attitude de rejet (sinon d'agacement). Ainsi si je dis :

J'aime bien les vieilles horloges et autres montres de poche, mais je n'aime pas les maths

Il y a de fortes chances pour qu'un éclat de rire prenne place chez le lecteur. En effet, le domaine couvert ici est celui des mécaniques en horlogerie, et la notion de

« maths » n'en fait pas partie. Notons que c'est tout de même grâce à ce mécanisme métalinguistique que fonctionne l'humour. Maintenant, si je dis :

Les femmes sont autorisées à se baigner, mais pas les filles

la préservation du domaine est certes respectée, cependant « femmes » et « filles » font partie d'un même domaine (ou praxème), et sont relativement à un même niveau de hiérarchie signifiante. Ce qui pose un problème d'entendement. Lorsqu'Antoine dit :

Il était mon ami, fidèle et juste ; mais Brutus dit qu'il était ambitieux

On crée un domaine avec deux niveaux de hiérarchie signifiante.

– L'un est porteur de traits positifs : + [grandeur d'âme, sensibilité, humanisme],

– L'autre de traits négatifs : – [grandeur d'âme, sensibilité, humanisme].

Procédant ainsi, Shakespeare réussit à créer chez l'auditeur/lecteur une sorte de télescopage à l'intérieur d'un même domaine qui se dit en antonymie. Une sorte de mise en écho antonymique à partir d'un même signifiant. Ainsi « César » se détache-t-il de la notion d'« ambition » et restaure un domaine autre. Vont contribuer à cet objectif deux autres types d'outils métalinguistiques : « être » et « avoir ». Précisons, au risque de paraître caricatural, qu'« être » permet de dégager un ou plusieurs traits identifiant un terme. À la limite, c'est un parfait outil de tautologie. « Avoir », de son côté, est un outil de localisation en ce sens qu'il détermine une relation de repérage sans laquelle l'identification serait hasardeuse. Partant de là, « avoir » constitue un excellent outil d'emboîtement/déboîtement praxémique (d'où l'effet de sens de « possession » qu'il induit). Souvent le terme repéré ou localisé constitue une information nouvelle ou rhématique. Ainsi :

Pierre a un chapeau

Est un énoncé qui localise « chapeau » par rapport à « Pierre » qui est un connu (puisqu'on le désigne par son prénom). L'opération de localisation (c'est-à-dire « d'entame discursive », comme dirait G. Guillaume) discursive la plus utilisée se matérialise souvent (en français) sous la forme d'« il y a ». Ainsi si « Pierre » n'avait pas été connu, on aurait eu recours à quelque chose comme

Il y a quelqu'un qui s'appelle Pierre, et bien il a un chapeau

Essayons de nous faire une idée du rôle de l'opérateur « avoir » dans la fameuse tirade d'Antoine.

(7) Et César l'a gravement expié.

(15) Il a ramené à Rome nombre de captifs

(16) Dont les rançons ont rempli les coffres publics :

- (18) Quand le pauvre *a gém*, César *a pleuré* :
- (23) Je lui *ai* trois fois présenté une couronne royale,
- (24) Qu'il *a refusé* trois fois :
- (29) Vous *l'avez* tous aimé naguère, et non sans motif ;
- (31) Oh jugement tu *as fui* chez les bêtes *brutes*,
- (32) Et les hommes *ont perdu* leur raison. Excusez-moi ;
- (36) Hier encore, la parole de César *aurait pu*
- (47) Je *l'ai trouvé* dans son cabinet ;

Le mécanisme induit par « avoir » permet d'enrichir nos connaissances d'un terme donné, c'est-à-dire opérer des emboîtements ou déboîtements. L'élément qu'« avoir » localise par rapport à un *repère*, *i.e.* le *repéré*, constitue une information nouvelle qui vient s'ajouter à ce que nous savions auparavant. Les auditeurs de Jules César, arrivés à l'acte III scène II, découvrent des propriétés de « César », jusque-là bien cachées :

- Les captifs ramenés à Rome l'ont été par César
- Les coffres publics remplis par l'argent des rançons l'ont été par César
- Les larmes versées sur le sort des pauvres l'ont été par César
- La couronne royale refusée l'a été par César

Le travail silencieux qu'effectue « avoir » se trouvera relayé par le non moins silencieux travail des connecteurs discursifs, qui, eux, assurent la cohésion et la cohérence textuelles. Le travail effectué par « être », pour ce qui le concerne, va consister essentiellement à présenter comme tautologique une relation privilégiée : « César était ambitieux ». Seulement Shakespeare ne s'en arrête pas là, il inscrit cette relation d'identification à l'intérieur d'une autre, englobante : « Brutus dit que + PROP ». L'intérêt du recours à la forme *Discours Rapporté*, c'est qu'elle opère un retournement thématique en cela que l'objet du « dit » est toujours repéré par rapport à « Brutus ». Ainsi finit-on par obtenir une relation où « Brutus » est le repère et « César + était + ambitieux », le repéré. Une opération de type « avoir », en somme. Au bout de quelques itérations « Brutus dit que X » finit par asseoir X comme un trait venant s'ajouter à ceux que nous connaissions auparavant. De la sorte, « César + était + ambitieux » devient un terme à localiser par rapport à « Brutus »... et non pas « César ». En sortie c'est quelque chose comme « Brutus est ambitieux », d'un côté ; et, de l'autre, « César était un ami, fidèle et juste », que

l'auditeur construit. La valeur prototypique de départ est ainsi largement dépassée ; voire marginalisée.

L'intérêt de notre illustration shakespearienne consiste en cette mise en exergue de ce mouvement d'emboîtement — déboîtement où un ensemble de traits (ou sèmes) une fois versés dans le discours, font l'objet de réglages. Ces derniers, à l'instar des mouvements propres aux sabliers, procèdent à des vidage et remplissage – voire substitutions – du programme de sens. Cependant il fallait bien partir d'un état zéro, un état à partir duquel le discours prend le relais. L'état zéro, c'est ce que l'on nomme le prototype. Par contre l'état final – bien qu'il conserve traces de l'état initial ainsi que des mouvements que lui fait subir le parcours discursif – identifie le signifiant à un autre signifié. Le signe paraît donc illusoire, certes, mais d'une illusion nécessaire. Tel est le paradoxe auquel F. de Saussure avait dû s'attaquer¹⁰ et que les travaux contemporains en sémantique¹¹ tentent de clarifier. Ajoutons qu'autant ces questions sont décisives lorsqu'il s'agit du traitement humain des langues naturelles, autant elles se posent différemment pour le traitement automatique¹². La notion « ambitieux », dans notre tirade shakespearienne, se vide bien de sa substance jusqu'à se désolidariser de son référent discursif initial (César). Cela l'humain le reconstruit bien. Et l'équation « César = ambitieux » se transforme en « César = ami fidèle et juste » ; malgré la stabilité des signifiants. Pour le traitement automatique, les procédures de reconnaissance des chaînes de caractères se basent exclusivement sur la transcription (graphique ou phonétique) du signifiant. Par conséquent, le dictionnaire rattaché à la procédure d'analyse automatique donnera, pour « ambitieux », la même liste de traits. Au début de la tirade, comme à la fin de celle-ci. Car, en définitive, seul l'humain est en mesure d'accéder au niveau des opérations métapraxémiques et, si l'on croit les tenants du programme minimaliste (cf. Chomsky 2000), l'humain possède un potentiel génétique et biologique qui le prédispose à cela. L'ordinateur pas¹³. En tout cas pas aussi longtemps qu'il ne rêvera pas.

BIBLIOGRAPHIE

- Adamczewski (H.), 1991 : *Le Français déchiffré*, Paris, A. Colin.
 Chomsky (N.), 2000 : *New Horizons in the Study of Language and Mind*, Cambridge University Press.
 Culioli (A.), 1990 : *Pour une linguistique de l'énonciation* tomes I & II, Paris, Ophrys.
 Culioli (A.), 1999 : *Pour une linguistique de l'énonciation* tome III, Paris, Ophrys.
 Delbecq (N.), édité, 2002 : *Linguistique cognitive : comprendre comment fonctionne le langage*, Bruxelles, Duculot.

¹⁰ Sur ces grandes questions, on lira avec beaucoup d'intérêt l'ouvrage de Saussure édité par S. Bouquet et R. Engler (2002).

¹¹ Les propos de O. Ducrot (2002) à ce sujet ont participé à donner de l'assurance à notre problématique.

¹² Les travaux du laboratoire de linguistique informatique de Paris XIII, sont à cet effet très explicites.

¹³ Insistons, tout de même, pour rappeler que l'ordinateur n'est sollicité que pour *restituer* des informations (dans une même langue ou dans une langue différente) ; l'humain, lui, *interprète* – malgré lui ? – les informations qu'il reçoit.

- De Saussure (F.), 2002 : *Écrits de linguistique générale*. Texte établi et édité par Bouquet (S.) et Engler (R.), Paris, Gallimard.
- Ducrot (O.), 2002 : « Le sens », dans *Le Cerveau, le Langage, le Sens*, vol. V, Paris, Odile Jacob, pp. 271-286.
- Elimam (A.), 2003 : *Le Maghribi, alias « ed-darija »*, Oran, Dar El-Gharb.
- Lafont (R.), 1978 : *Le Travail et la Langue*, Paris, Flammarion.
- Lerat (P.), 1995 : *Les Langues spécialisées*, Paris, PUF.
- Pollock (J. Y.), 1997 : *Langage et Cognition*, Paris, PUF.

LE CONTEXTE : FILTRE OU MEMBRANE ?

Christine Durieux

CRISCO (FRE 2805 CNRS) et MODESCOS MRSH

Université de Caen

Caen, France

christine.durieux@unicaen.fr

***Résumé :** Quand le contexte joue le rôle de filtre, il laisse passer des potentialités de significations non pertinentes pour ne retenir que celle qui se prête à la construction d'un sens. Quand le contexte joue le rôle d'une membrane, il permet des échanges avec le milieu extérieur. Dans le cadre théorique de la complexité, le contexte n'est pas un donné stable, mais un construit qui certes influe sur la valeur des mots, mais se trouve aussi constamment modifié par la valeur que lui-même confère aux mots.*

***Mots-clés :** complexité, construction du sens, contexte*

1. INTRODUCTION

En traductologie, la théorie interprétative de la traduction a depuis longtemps mis en évidence le fait que le sens n'est pas dans les mots et que « le sens d'un énoncé n'est pas la somme de la signification des mots successifs pris isolément qui le composent » (Durieux 1988 : 42). Les mots, somme toute simples chaînes graphiques conventionnelles, ne livrent une signification qu'à l'issue du déroulement d'un mécanisme de perception relayé par tout un processus d'interprétation qui met en jeu bien plus que la seule reconnaissance des mots et leur mise en correspondance avec leur définition connue telle que celle que pourrait donner un dictionnaire. Ainsi, c'est la prise en compte d'un large contexte verbal et cognitif qui permet de comprendre qu'une voiture verte n'est pas nécessairement une voiture dont la peinture est de couleur verte, mais peut être un véhicule peu polluant, équipé d'un moteur hybride, par exemple. Dans ce cas, la mention de la couleur verte est déconnectée de sa signification conventionnelle correspondant à une longueur d'onde lumineuse approximative de 500 à 560 nm. La filiation est très indirecte et passe par toute une série de relations symboliques et analogiques.

2. LE CONTEXTE COMME FILTRE

La réfutation de la thèse de l'autonomie du sens conduit logiquement à convoquer la thèse de la dépendance contextuelle. Selon ce principe, les unités linguistiques composant l'énoncé font sens en fonction du contexte. Le contexte permet de lever les ambiguïtés et d'actualiser des acceptions et des significations virtuelles au détriment de toutes les autres possibles. Ainsi, le contexte joue le rôle de filtre,

laissant passer des potentialités de significations non pertinentes pour ne retenir que celles qui se prêtent à la construction d'un sens.

Pour illustrer cette approche, nous avons choisi un article de vulgarisation récent intitulé « Les miracles du plasma ». Ce choix se justifie par le fait qu'à la lecture du titre, le lecteur hésite entre au moins deux attentes : s'agit-il d'un texte relevant du domaine de la biologie qui traite du plasma sanguin, ou de l'audiovisuel avec les écrans à plasma ? Certes, on peut faire des miracles avec le plasma sanguin, c'est-à-dire la partie liquide du sang, administré en perfusion à des malades. Par ailleurs, à l'heure actuelle, avec la rapidité des progrès technologiques, les tubes cathodiques sont de plus en plus supplantés par des écrans plats dont certains sont l'application de la technologie des plasmas. La seule lecture du titre ne permet donc pas d'orienter l'attente du lecteur et la nécessité d'acquérir des éléments de contexte apparaît d'emblée. Dans cet exemple précis, le contexte situationnel n'apporte pas d'éclaircissement.

LE MIRACLE DES PLASMAS

Injectés autour d'un avion, les plasmas isolent l'appareil de l'air. Rendu plus fluide, l'écoulement aérodynamique oppose alors une moindre résistance.

Avec l'avènement des plasmas – une soupe électromagnétique, où ions et électrons sont mélangés et créent un milieu globalement neutre – l'aéronautique est à la veille de connaître un bouleversement.

Le halo lumineux légèrement bleu violacé dans lequel va évoluer l'appareil pendant le vol est un phénomène tout à fait normal, volontairement provoqué par un champ de plasma pour affranchir l'avion de la résistance de l'air.

L'écoulement de l'air autour de la voilure s'accélère, et la couche-limite n'est plus tourbillonnaire. Elle en épouse le profil sans se décoller.

En effet, l'article est publié dans la rubrique Sciences d'un hebdomadaire grand public, ce qui se trouve confirmé par la présence du terme technique/scientifique *plasma* dans le titre, et le domaine annoncé est l'aéronautique. *A priori*, on ne voit guère de relation entre le plasma, qu'il soit sanguin ou gazeux, et l'aéronautique. Toutefois, le chapeau permet de réaliser un premier tri dans les significations possibles du terme *plasma*¹.

« Injectés autour d'un avion, les plasmas isolent l'appareil de l'air. Rendu plus fluide, l'écoulement aérodynamique oppose alors une moindre résistance. »

¹ *Le Grand Robert de la langue française* (2001) propose trois significations : apparue en 1752, variété verte de calcédoine, minéral apparenté au quartz ; apparue en 1845, partie liquide du sang ; apparue en 1925, gaz ionisé électriquement neutre, porté à haute température, riche en ions et en électrons libres.

La présence dans le contexte immédiat des mots/termes – avion, appareil, air, écoulement aérodynamique, résistance – constitue un ensemble isotopique qui conduit à exclure les potentialités de significations du terme *plasma*, telles que variété de calcédoine vert foncé ou partie liquide du sang, pour ne retenir que le quatrième état de la matière qui est celui de gaz ionisé. Le contexte verbal joue donc un rôle de filtre. Il filtre la polysémie et contribue à actualiser une acception du terme plasma au détriment de toutes les autres possibilités². On remarque, toutefois, que l'exclusion des autres acceptions connues et répertoriées, pour fait d'in vraisemblance, n'est possible que si le lecteur se fonde sur ses connaissances thématiques. De fait, sans l'apport de connaissances techniques, la désambiguïsation ne peut se faire. Ce n'est pas la seule présence des unités linguistiques dans le contexte verbal qui fait fonction de filtre, mais leur interprétation à la lumière d'un savoir encyclopédique. Or, il ne suffit pas toujours au lecteur de rappeler de sa mémoire à long terme le savoir encyclopédique qu'il y a stocké pour pouvoir construire un sens pertinent. Il arrive qu'il ne dispose pas des connaissances nécessaires et qu'il doive les chercher. À cet effet, les sources documentaires sont de deux natures : intratextuelles et extra-textuelles. Certes, il est courant de rechercher dans un dictionnaire monolingue et des encyclopédies des définitions ou des explications de termes recouvrant des notions ignorées ou mal connues du lecteur. Or, bien souvent, une exploration intratextuelle fournit la solution. Ainsi, le plasma est défini plus loin dans l'article étudié comme une « soupe électromagnétique ». Cette collocation hardie est révélatrice d'une métaphore. En effet, il y a rupture de domaine entre le substantif (nourriture) et l'adjectif le qualifiant (électronique). Là encore, le contexte joue le rôle de filtre en laissant passer les sèmes non actualisés du mot soupe – « bouillon épaissi par des tranches de pain ou des aliments solides », « nourriture simple et substantielle », « repas rustique et populaire », etc. – pour ne retenir que la valeur de mélange d'ingrédients dans des proportions non déterminées : « où ions et électrons sont mélangés et créent un milieu globalement neutre » poursuit l'article.

Avant de la critiquer, J. Moeschler énonce ainsi la thèse du filtrage contextuel : « Le sens d'une expression en usage est un sous-ensemble de ses sens possibles. Le contexte fonctionne comme un filtre qui sélectionne la valeur contextuellement appropriée de l'expression. » (Moeschler 1996 : 58). Toutefois, ce raisonnement présente une grande faiblesse. En effet, s'il est applicable sur un plan pragmatique simple, il se heurte à une contradiction épistémologique. De fait, il faudrait que l'on puisse constituer et cerner un ensemble fermé de toutes les significations possibles d'un mot et que l'on admette que le contexte contribue à les invalider

² Pour le terme *plasma*, on parle de polysémie parce que les trois acceptions figurent dans les dictionnaires sous la même entrée-vedette et sont données comme procédant d'une étymologie commune. Le terme, tel qu'il existe en français contemporain, serait emprunté au grec *plasma*, nom d'action dérivé du verbe *plassein* qui signifie façonner ou modeler. La dérivation s'est poursuivie en français avec l'adjectif et le substantif *plastique*. En revanche, selon la théorie du prototype, on parlerait dans ce cas d'homonymie parce qu'il paraît difficile de maintenir un noyau prototypique commun à ces trois désignations. De plus, curieusement, les trois réalités désignées relèvent d'états différents de la matière : solide en minéralogie, liquide en biologie et gazeux en physique.

toutes sauf une, c'est-à-dire la seule pertinente *hic et nunc*. Or, d'une part, on ne saurait construire un tel ensemble borné : la langue est vivante et offre en permanence la possibilité de nouvelles combinatoires et de nouvelles significations ; l'ensemble des potentialités est donc un ensemble ouvert qui ne peut faire l'objet d'un tel traitement. D'autre part, on peut s'interroger sur la manière dont se ferait cette sélection. Quelles seraient les caractéristiques des significations ainsi élues ? Y aurait-il des récurrences qui permettraient de prédire le résultat de la sélection opérée ?

3. LE CONTEXTE COMME MEMBRANE

À la différence du filtre qui ne fonctionne que de façon unidirectionnelle, la membrane fonctionne de façon bidirectionnelle : sa fonction est d'assurer les échanges avec son environnement.

« Le halo lumineux légèrement bleu violacé dans lequel va évoluer l'appareil pendant le vol est un phénomène tout à fait normal, volontairement provoqué par un champ de plasma pour affranchir l'avion de la résistance de l'air ».

La collocation *champ de plasma* surprend le lecteur non spécialiste [rappelons qu'il s'agit d'un article paru dans un magazine hebdomadaire grand public] pour qui un champ est d'abord un terrain cultivé et dont l'attente spontanée est donc la dénomination d'une plante en complément du mot champ. Toutefois, une certaine polysémie est couramment attachée à ce mot/terme : à côté de champ de blé ou de maïs, il y a bien champ de bataille, mais aussi champ visuel ou champ magnétique. Ainsi, le plasma constituerait-il un milieu assimilable à un champ électromagnétique. Pour autant, là encore, les seules composantes linguistiques ne permettent pas l'appréhension du sens de cet énoncé. D'une part, le rôle de filtre assuré par le contexte verbal reste subordonné à l'intervention d'un contexte cognitif. D'autre part, il y a une relation manquante : quel lien établir entre la création d'un champ électromagnétique et la diminution voire la suppression de la résistance opposée par l'air ? On sait que tout véhicule en déplacement rencontre une résistance à la pénétration dans l'air. Pour limiter cette résistance, la réponse a jusqu'à présent été apportée par les concepteurs qui ont cherché à dessiner des profils de véhicules – voitures, bateaux, avions – toujours plus aérodynamiques pour améliorer la pénétration dans l'air et faciliter l'écoulement de l'air le long des structures. Dans l'article étudié, justement il ne s'agit pas d'architecture profilée mais du rôle d'un champ de plasma. Ici, le contexte joue un rôle de membrane dont la fonction est d'assurer les échanges avec son environnement. De fait, les connaissances acquises au fur et à mesure de la lecture du texte confèrent aux unités linguistiques composant le texte une signification propre de nature à permettre la construction du sens cohérent. Par exemple, s'étant fait la représentation d'un champ de plasma comme d'une couche gazeuse enveloppant l'avion, servant en quelque sorte de lubrifiant pour faciliter sa pénétration dans l'air, le lecteur est mieux à même de comprendre « lorsque le plasma est produit, l'écoulement de l'air autour de la voilure s'accélère ». À ce stade de la construction

du sens, la voileure ne devrait susciter alors aucun conflit polysémique. Les acceptions telles que l'ensemble des voiles d'un navire ou l'état d'une roue de bicyclette voilée n'entrent pas en concurrence ; elles sont d'emblée occultées au profit de l'ensemble des surfaces portantes d'un avion. Il n'y a pas de calcul du sens, mais un sens qui s'impose de lui-même en fonction du contexte verbal comme déclencheur de l'activation d'un vaste réseau de contexte cognitif. « L'un de leurs sens [des mots] exclut les autres et vient s'imposer dans l'énoncé ; une fois encore, c'est le tout qui contribue à donner sens à la partie, laquelle contribue à donner sens au tout. » (Morin 1991 : 170). En réalité, le sens présente une saillance qui fait qu'il s'impose comme une forme se détache d'un fond (théorie de la Gestalt). Autrement dit, au principe instructionnel qui veut que les instructions de nature à permettre la construction du sens soient encodées dans la matière linguistique, nous opposons le principe inférentiel selon lequel la construction du sens se fait selon un modèle cognitif axé sur le déroulement d'un raisonnement logique fondé sémantiquement (Durieux 2005).

4. LE CONTEXTE COMME CONSTRUCTION À GÉOMÉTRIE VARIABLE

Au fur et à mesure de la lecture/compréhension, le contexte s'étoffe. Le contexte n'est donc pas un donné stable, préexistant à l'acte de lecture/compréhension ; c'est un construit qui certes influe sur la valeur des mots, mais se trouve aussi constamment modifié par la valeur que lui-même confère aux mots. De plus, ce faisant, il contribue à modifier les croyances du lecteur et son savoir qui, à son tour, va être projeté dans l'opération d'interprétation dans le cadre de l'exercice de lecture/compréhension de la suite du texte. On relève donc des interactions constantes entre le texte, objet de l'opération de lecture/compréhension, et plusieurs niveaux de contexte.

Le co-texte immédiat permet une désambiguïsation spontanée. C'est le cas, par exemple, d'une collocation telle que la *soupe électromagnétique* mentionnée plus haut. On remarque que l'adjectif reflète une réalité alors que le substantif présente une valeur métaphorique. Cela n'empêche qu'on pourrait imaginer un roman dans lequel un personnage mettrait un électroaimant dans la soupière (!), auquel cas la collocation ne ferait pas l'objet d'un phénomène tropologique. En revanche, dans une collocation comme le *plasma embarqué*, le trope – ici une métonymie – porte sur l'adjectif, le substantif dénotant une réalité correspondant à une définition répertoriée.

Le contexte verbal, support de toute une isotopie, contribue à préciser la signification de tout mot/terme polysémique. Dans l'article étudié, c'est le cas notamment de *plasma*, mais aussi de *voilure*. On peut remarquer avec un certain intérêt que les mots/termes cités en exemple ne sont pas des unités du vocabulaire courant dont on connaît bien le caractère normalement polysémique, mais précisément des termes techniques réputés être monoréférentiels. Ce qui conduit à considérer que tous les mots d'une langue sont polysémiques ou susceptibles de

l'être, de même que tous les termes, si spécifiques et monoréférentiels soient-ils, se prêtent à des emplois métaphoriques³.

Le contexte cognitif englobe toutes les connaissances acquises par le lecteur, qu'il a stockées dans sa mémoire à long terme et qu'il peut mobiliser le moment venu pour construire un sens à la lecture d'un texte. C'est son savoir de type encyclopédique qui permet au lecteur de comprendre l'énoncé : « l'écoulement de l'air autour de la voilure s'accélère ». Comprendre cet énoncé revient à se faire une représentation mentale qui inverse les facteurs linguistiques. En effet, on sait ce que signifie l'accélération d'un moteur, en automobile, par exemple : le véhicule en mouvement se déplace plus vite. Mais dans le cas de l'écoulement de l'air évoqué dans l'article étudié, le véhicule en mouvement est l'avion ; pour reprendre une métaphore lexicalisée illustrative, il fend l'air ; mais l'air lui-même ne peut s'écouler autour de la structure de l'avion tant que ce dernier ne l'a pas « fendu », la notion d'accélération n'est donc pas la même que dans le cas de l'automobiliste qui accélère. À cet égard, il n'y a aucune instruction ni aucun indice linguistique qui permette de construire le sens, c'est bien à sa capacité de raisonnement exploitant un savoir de nature encyclopédique que le lecteur a recours pour se faire une représentation mentale cohérente, c'est-à-dire pour comprendre cet énoncé.

Le contexte hypertextuel est une construction en réseau, qui se modifie constamment au fil des parcours interprétatifs, chaque fois différents. À chaque instant de la lecture, une nouvelle association, une nouvelle interprétation vient modifier le contexte, l'accroître, le réorienter. « L'opération élémentaire de l'activité interprétative est l'association ; donner du sens à un texte quelconque revient à le relier, le connecter à d'autres textes, et donc à construire un hypertexte. » (Lévy 1990 : 80). Par exemple, l'énoncé « l'écoulement de l'air autour de la voilure s'accélère, et la couche-limite n'est plus tourbillonnaire. Elle en épouse le profil sans se décoller » fait appel à des notions développées dans d'autres textes spécialisés. La notion de mouvement tourbillonnaire renvoie à la théorie du chaos et aux théories de la complexité, et c'est précisément dans le cadre de ces théories que le sens de l'énoncé ci-dessus peut, et doit, être appréhendé.

³ *Le Grand Robert de la langue française* (2001, tome V : 772) cite un extrait de la correspondance de Jean-Richard Bloch avec Romain Rolland dans lequel le terme *plasma* est employé de façon métaphorique : « Je travaille... peinant des journées pour quatre pages utiles, comprimant l'intention jusqu'au plasma essentiel, et voyant néanmoins le manuscrit enfler ».

De même, un terme scientifique très spécialisé comme *tsunami*, emprunté au japonais, est présent dans la langue spécialisée de la géographie en français depuis 1927 et est resté jusque très récemment ignoré du grand public. Or, sous l'effet du matraquage des médias, il a fait irruption dans le vocabulaire courant avec la catastrophe qui a dévasté l'Asie du Sud-Est en décembre 2004. Désormais, ce terme, qui désigne pourtant un phénomène tellurique très spécifique, est couramment employé de façon métaphorique, notamment dans la grande presse où on peut lire : ... après le tsunami politique du 29 mai dernier (non majoritaire au référendum en France sur le projet de texte de la Constitution européenne), ou encore... la privatisation d'EDF va provoquer un tsunami boursier.

5. CONCLUSION

S'il est utile sur le plan pédagogique de dissocier les différents niveaux de contexte, il n'en reste pas moins que, comme toute classification, cette catégorisation est réductrice. En effet, tous les niveaux sont interdépendants. Le contexte étroit ou le plus large contexte verbal ne peut jouer son rôle de filtre ou de membrane que s'il peut être associé à un contexte cognitif solide chez le lecteur. En outre, ce contexte cognitif, en perpétuelle mouvance sous l'effet de l'apprentissage et de l'expérience de chaque instant, se fond dans un plus vaste contexte hypertextuel qui, poussé à l'extrême, couvre l'ensemble du savoir humain.

Cette catégorisation s'intègre dans une approche pédagogique pragmatique, qualifiable à la limite d'utilitaire. Bien sûr, la question de la relation entre mot/terme et contexte pourrait être abordée d'une tout autre manière, notamment en tentant de répondre à la question sous-jacente à toute cette réflexion : qu'entend-on par contexte ? À ce moment-là, la problématique s'inverse, on s'intéresse alors au contexte en tant que terme à définir, avec ses différentes acceptions... en fonction du contexte dans lequel il est employé ; et la boucle est bouclée.

En tout état de cause, il y a interdépendance entre le mot/terme et la situation de communication dans laquelle il est employé. Sans adhérer aveuglément à l'affirmation de Wittgenstein selon laquelle « les mots n'ont pas de sens, ils n'ont que des emplois », force est de constater que ce point de vue se trouve bien confirmé dans la réalité, avec l'extension illimitée de la polysémie, la métaphorisation de termes techniques très spécialisés et le vocabulaire extrêmement limité du parler des banlieues qui utilise les mêmes mots pour désigner des objets ou des notions très diverses.

BIBLIOGRAPHIE

- Durieux (C.), 1988 : *Fondement didactique de la traduction technique*, Didier Érudition, Paris.
- Durieux (Cr.), 2005 : « La traduction, illustration d'un processus complexe », dans *La Complexité* (R. Hérin et B. Cadet, dir.), Cahiers de la Maison de la recherche en sciences humaines, Caen, pp. 179-192.
- Lévy (P.), 1990 : *Les Technologies de l'intelligence*, La Découverte, Paris.
- Moeschler (J.), 1996 : *Théorie pragmatique et pragmatique conversationnelle*, A. Colin, Paris.
- Morin (E.), 1991 : *La Méthode 4. Les idées*, Seuil, Paris.

XML/XSL(T) POUR UN TRAITEMENT UNIFIÉ DE *CORPUS* MULTILINGUES ET DES RÈGLES DE TRADUCTION CORRESPONDANTES

Pascaline Merten

Institut supérieur de traducteurs et interprètes (ISTI)

Haute École de Bruxelles (HEB)

Bruxelles, Belgique

pmerten@heb.be

Résumé : *La recherche linguistique doit se fonder sur de nombreux exemples. Ces exemples sont habituellement tirés de lectures ou de corpus. Si la recherche est comparative et porte sur différentes langues, les corpus et les exemples doivent en plus couvrir plusieurs langues, être parallèles ou alignés.*

Cet article part du problème linguistique que l'on souhaite étudier et propose différentes méthodes visant à la constitution de corpus de textes monolingues et multilingues ainsi qu'à l'extraction des exemples qui fonderont la description linguistique. L'emploi de langages standards de la famille XML et le recours à des normes facilitent les manipulations informatiques et favorisent la diffusion des corpus et des procédures. En particulier, le langage XSL pourra être utilisé tant pour représenter des corpus de manière lisible que pour effectuer de véritables traitements et ainsi tester des hypothèses linguistiques.

Mots-clés : *XML, XSL, XSLT, TEI, XCES, corpus, corpus multilingues, linguistique, alignement, étiquetage*

1. INTRODUCTION

Le recours aux *corpus* est devenu une pratique fréquente. Internet, pour commencer, offre une mine d'exemples au linguiste soucieux de cerner l'emploi des mots, des termes, des structures. Divers sites proposent des ressources lexicographiques ou documentaires incontournables.

Il n'empêche que pour des besoins spécifiques, ces ressources libres ne suffisent pas toujours et il faudra construire ses propres ressources, monolingues ou multilingues. La constitution de *corpus* peut être un véritable parcours du combattant, surtout s'il s'agit de *corpus* parallèles. De surcroît, bien des recherches linguistiques ne peuvent se satisfaire de *corpus* bruts, mais doivent accéder à la forme lemmatisée ou à la catégorie morphosyntaxique. En pareil cas, il faut catégoriser le *corpus* ou faire appel à des *corpus* catégorisés. Et qui dit *corpus* catégorisé dit recherche sur des expressions régulières. Peu de *corpus* publics sont catégorisés, et les langages d'interrogation utilisés pour les consulter sont souvent en deçà des expressions régulières.

Il n'empêche qu'une fois ce travail fait, il convient de représenter les *corpus* de textes et d'exemples de façon normalisée, de manière à ce que ces ressources soient réexploitables par d'autres ou à d'autres fins.

2. LE PROBLÈME LINGUISTIQUE

Notre questionnement sur les *corpus* et leur représentation est parti d'un questionnement linguistique, d'une recherche sur un problème linguistique précis : qu'est-ce qui, en français, régit l'ordre des qualifiants ; et s'il faut traduire ce genre de structure, l'ordre des mots est-il fondé sur les mêmes critères dans toutes les langues ?

2.1 De l'adjectif au qualifiant

Le terme d'*adjectif* porte sur la nature des mots : *bleu*, *benzénique*, *supplémentaire*, *halogéné* sont des adjectifs. Le terme de *qualifiant* désigne une fonction syntaxique : celle des déterminants du nom qui le délimitent en extension (par opposition aux quantifiants qui donnent une indication d'extensité) (Wilmet 2003 : 121)¹.

Dans l'exemple

[1] Les liaisons chimiques sont les forces qui maintiennent les atomes entre eux dans des arrangements spécifiques qui forment les molécules,

l'adjectif *chimiques* qualifie le nom *liaisons*, il précise de quelles liaisons chimiques on parle, en termes ensemblistes, il circonscrit un sous-ensemble.

Le quantifiant *les* indique quels sont les éléments pointés de ce sous-ensemble, en l'occurrence tous.

Dans

[2] Le noyau contient des protons chargés positivement et des neutrons non chargés,

le quantifiant *des* pointe non sur la totalité des protons chargés positivement mais sur un certain nombre d'entre eux.

En passant de la nature à la fonction, on s'intéressera à tous les mots qui peuvent ainsi délimiter l'extension du nom :

le nom : le pont Mirabeau, une voiture tout-terrain, un ticket-repas ;

le complément du nom (CN) : le chien du berger, un chien de berger, un homme de loi, un homme de bien, le code de la route... ;

la proposition relative : l'homme qui rit, l'idée que tu partes ;

l'infinitif : l'idée de partir

¹ Dans son ouvrage de 2003, Wilmet emploie le terme de caractérisant. Le terme de qualifiant est celui qu'il emploie dans son *Précis de grammaire française* (à paraître). Nous avons déjà adopté sa nouvelle terminologie.

le nom propre : le roi Louis, la ville de Paris ;

l'adverbe : le toujours ministre ;

le pronom : le bien d'autrui ;

la phrase : le numéro je ne sais plus combien de la rue je ne sais plus comment.

2.2 Problématique en français

Le problème linguistique qui nous a intéressée est la séquence de ces qualifiants entre eux, ou *ordre relatif*, qu'il s'agisse d'une séquence d'adjectifs antéposés :

[3] une charmante petite vieille dame

ou d'une séquence d'adjectifs postposés :

[4] l'acide sulfonique aromatique

[5] des groupes caractéristiques nombreux

ou encore d'un mélange de classes morphosyntaxiques :

[6] une compagnie de navigation marseillaise

[7] de l'eau qui pique à l'orange

[8] la chaîne de télévision italienne Canale 5

[9] le logiciel résidant en mémoire *McAfee*.

L'observation de tels exemples appelle divers commentaires. Dans certains cas, deux séquences sont possibles, avec des différences de sens plus ou moins marquées : l'exemple [3] peut devenir

[3bis] une charmante vieille petite dame

et l'exemple [4] peut devenir

[4bis] l'acide aromatique sulfonique.

Dans d'autres cas, un des adjectifs peut changer de place absolue (antéposition/postposition). L'exemple [5] peut ainsi devenir

[5bis] de nombreux groupes caractéristiques.

Changer la place des noms propres dans les exemples [8] et [9], en revanche, conduit à des groupes incorrects – sauf à les considérer comme des appositions :

[8bis] *la chaîne Canale 5 de télévision italienne

[9bis] *le logiciel McAfee résidant en mémoire.

Mais, quoi qu'il en soit de la grammaticalité des chaînes, deux séquences différentes ont des sens différents, même si l'objet du monde reste le même : *une compagnie française de navigation* et *une compagnie de navigation française* désignent

peut-être la même compagnie, mais le signifié est différent. Ceci se marque aussi dans l'emploi des quantifiants :

[10] des globules rouges résistants

[11] une robe du soir rouge

[12] une robe rouge échanquée

[13] la robe échanquée rouge (et non la noire).

2.3. Confrontation avec d'autres langues

On sait que les langues romanes antéposent certains adjectifs, en postposent d'autres et que l'ordre majoritaire y est la postposition. On sait que les langues germaniques antéposent au contraire massivement les adjectifs². C'est pour l'anglais qu'on trouve la majorité des études sur la question de l'ordre relatif des « modificateurs » puisque cette langue antépose non seulement les adjectifs mais les noms en construction directe :

[14] *my dear old happy Belgian sheep dog.*

D'aucuns considèrent que l'ordre des adjectifs dans les langues romanes est fondamentalement l'antéposition (Cinque 1994 : 85). Cette position est difficilement tenable d'abord en diachronie puisque l'évolution va dans le sens de la postposition. En synchronie également, il y a des différences d'emploi et de sens liées à la place absolue et qui ne se laisse pas réduire : comment ramener le sens du *chapitre dernier* au sens du *dernier chapitre* ? Il serait plus simple de déduire le sens antéposé du sens postposé que l'inverse !

D'autres considèrent que la séquence des modificateurs dans les langues à postposition est l'image en miroir de cette séquence dans les langues à antéposition (Sal-koff 1999 : 120, Siewierska 1988 : 63). C'est sans doute vrai au niveau de l'organisation générale du syntagme (Rijkhoff 1988), en particulier pour ce qui est de la place des quantifiants, et pour la séquence générale des modificateurs d'un point de vue sémantique. Néanmoins, la traduction ne peut s'accommoder d'une vue aussi générale. Le choix du traducteur peut interférer, comme les différences morphosyntaxiques ou les différences stylistiques entre langues.

[15] la Cour internationale de justice = *Internationaal gerechtshof* = *International Court of justice*

[16] *Koninklijk Belgisch Comité* = Comité royal belge

[17] *un sistema generoso di tutele* = un système de protection généreux

[18] *la sostanza raccapricciante del romanzo* = l'affreuse substance du roman.

En [15], on observe que l'adjectif « géographique » vient s'intercaler entre le nom de l'institution et le complément qui en exprime la finalité. On aura en revanche le

² Voir néanmoins Teyssier (1968) pour une discussion sur les cas de postposition de l'adjectif en anglais.

tribunal pénal international. En néerlandais, la finalité de l'institution donne lieu à un nom composé et l'adjectif suit.

L'exemple [16] montre la tendance du français royal à précéder l'adjectif géographique alors que le néerlandais présente la tendance inverse.

Les exemples [17] et [18] témoignent de différences d'ordre absolu et relatif entre le français et l'italien.

3. LE POURQUOI DES *CORPUS*

La multiplication des qualifiants n'est pas extrêmement fréquente en français. Elle peut devenir parodique, ironique, bref, stylistiquement marquée, mais dans les langues de spécialité elle est tout à fait neutre. Si on peut figer certaines séquences en les considérant comme des termes, d'autres ne sont que partiellement figées : dans l'exemple [5], *groupe caractéristique* est lexicalisé, mais pas *groupe caractéristique nombreux*. Et de toute façon, que l'on fige certaines séquences peut se comprendre dans certains contextes, cela n'empêche pas d'examiner ces groupes sous un jour linguistique, pour découvrir comment se constituent les termes tant sur le plan de la syntaxe que sur celui de la sémantique.

Il faut donc de nombreux exemples pour dégager des généralités. On ne peut se contenter de faire quelques recherches sur Internet. Qu'y chercherait-on ? Il n'y a pas moyen de dire que l'on veut une suite de deux adjectifs ou une suite complétement du nom-adjectif. On peut juste chercher une séquence précise, une occurrence d'un adjectif. C'est utile, mais d'une utilité très limitée. Qui plus est, si l'on s'interroge sur l'emploi de *nombreux*, on devra chercher successivement *nombreux*, *nombreuse* et *nombreuses*. On peut s'estimer heureux de n'avoir que trois recherches à lancer, encore faut-il s'attendre à un bruit considérable au regard des structures étudiées.

Si de plus, on souhaite pouvoir établir des comparaisons avec d'autres langues, on devrait idéalement constituer des *corpus* bilingues. Soit, on partira de documents déjà traduits, soit on traduira soi-même. Le travail est de toute façon énorme.

Ajoutons que les *corpus* attendus sont des *corpus* d'exemples, dans notre cas de groupes nominaux. Pour les établir, on part de *corpus* de textes, monolingues, bilingues ou parallèles. La problématique n'est pas tout à fait la même. On peut avoir déjà un *corpus* d'exemples dans une langue tierce, et le traduire en français. On peut avoir un *corpus* de textes (parallèles ou non), et il faudra en extraire manuellement ou automatiquement les séquences pertinentes.

4. LA CONSTITUTION DES *CORPUS*

La question se pose alors de la constitution de ces *corpus*. La finalité est d'avoir des exemples en français, ainsi que des exemples bilingues. Il faudra faire une différence fondamentale selon qu'on constitue un *corpus* d'exemples, au départ dans une seule langue, ou un *corpus* textuel, lequel servira de base à une liste

d'exemples. Ainsi, on conservera, on transcrira les exemples rencontrés au hasard des lectures. Ils peuvent être uniquement en français ou bilingues. Par exemple, nous avons rassemblé des exemples de dénominations officielles, comme *Cinémathèque royale de Belgique* ou *Tribunal pénal international*, que nous avons repérés en français et dont nous avons ensuite recherché la traduction. D'autre part, on peut constituer un *corpus* de textes relevant d'un domaine. Les exemples qui illustrent la suite de notre propos sont tirés d'articles de journaux, d'ouvrage sur l'architecture, sur la chimie organique ainsi que d'interfaces de logiciels.

4.1. *Corpus* d'exemples

Les exemples collationnés peuvent l'être dans un simple document texte ou de traitement de texte. Ce qui importe lorsqu'on enregistre des exemples est de pouvoir les retrouver ; et ce qui nous intéressera est de pouvoir retrouver les exemples selon leur configuration syntaxique ou selon l'adjectif. Ainsi, il faudrait pouvoir retrouver les suites adjectif-adjectif-nom (deux adjectifs antéposés avant le nom) ou les suites adjectif-nom-adjectif (deux adjectifs répartis autour du nom) ou encore les séquences adjectif-nom-CN, nom-CN-adjectif et nom-adjectif-CN. Ceci peut se réaliser de différentes façons. On peut coder les exemples en ajoutant une annotation morphosyntaxique à chacun des mots de l'exemple ; on peut aussi faire appel à un SGBD de type *Access* où différents champs permettront de coder les informations pertinentes. Nous y reviendrons dans la section « représentation des *corpus* ».

La traduction de tels exemples peut se faire manuellement ou par recherche d'équivalents. Ainsi, la recherche des traductions de dénominations officielles s'est faite par recherche sur les sites des organisations concernées ou via la base de données terminologiques européenne Eurodicautom. Nous préconisons de le faire dans une *mémoire de traduction* qui offre un format d'import-export XML. Mais, moyennant une conversion informatique, il y a moyen d'enregistrer ces équivalences dans une base de données relationnelle ou dans un tableau.

4.2. *Corpus* de textes

Les *corpus* de textes peuvent également être monolingues ou bilingues. On mettra au nombre des *corpus* monolingues, la base de données textuelle *Frantext*³, dont la version catégorisée constitue un trésor offert à la curiosité du linguiste. Mais ce n'est bien sûr pas tout de disposer de documents pertinents, encore faut-il y retrouver les exemples intéressants. Si l'on se contente de remplacer le document traditionnel par un document électronique pour le lire à l'écran plutôt que sur papier, on ne sera guère avancé. La récupération au format numérique de documents peut se faire par exploitation des fichiers ou par numérisation du document papier. On peut par exemple récupérer des articles français et italien sur les sites du *Monde diplomatique* français et italien.

³ L'adresse est <http://www.frantext.fr/>. *Frantext* n'est pas en libre accès, mais fonctionne par abonnement aux institutions.

Si l'on dispose des documents dans deux langues, il sera intéressant de les aligner.

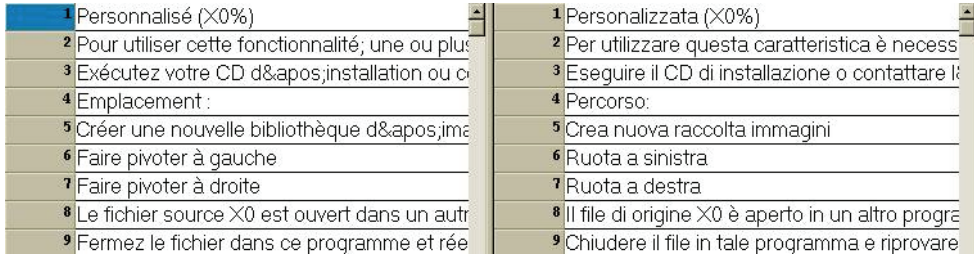


Figure 1 : alignement

L'alignement consiste, à partir de deux fichiers qui constituent l'équivalent linguistique l'un de l'autre, en l'établissement des équivalences entre segments correspondants.

L'alignement est une tâche qui demande une connaissance des langues au moins passive. C'est une tâche fastidieuse, coûteuse en temps parce que l'alignement automatique doit systématiquement être vérifié par l'utilisateur humain. En effet, les outils disponibles actuellement sont à base statistique et non linguistique⁴. Lorsqu'un segment dans une langue n'a pas d'équivalent dans l'autre langue, lorsque la différence de longueur entre deux segments est trop importante, lorsqu'il n'y a pas de correspondance 1-1, l'aligneur est mis en défaut : l'utilisateur doit corriger et relancer l'alignement. La tâche peut être fastidieuse, mais le résultat est à la hauteur. Les outils d'alignement proposent en effet un format d'import-export, le format TMX, format d'échange développé par la Lisa⁵ et servant de format d'échange entre mémoires de traduction. Ainsi, l'alignement obtenu peut être récupéré dans n'importe quelle mémoire de traduction, mais aussi utilisé tel quel comme on le verra ci-après.

5. LA REPRÉSENTATION DES CORPUS

Le document s'est fait numérique ; il trouve son pendant informatique dans la notion de fichier. Au texte du document s'ajoute une interprétation, qui sera, au minimum, la ponctuation, l'indication des divisions du texte, des titres, des éléments mis en évidence, des références et des renvois... C'est ce qu'on appelle le balisage⁶. Ce balisage peut être codé de manière spécifique à une application : on parlera de format propriétaire, comme c'est le cas dans les traitements de texte commerciaux. Ce balisage peut respecter une norme, de manière à être connu de tous et utilisable par tous. C'est ainsi que, pour représenter des documents et des *corpus*,

⁴ Exception faite de *Similis*, développé par Lingua et Machina (<http://www.lingua-et-machina.com/>).

⁵ Localization Industry Standards Association : <http://www.lisa.org/>. Le format TMX a été développé dès 1998 et constitue une des premières applications XML.

⁶ Pour plus de détails sur les notions de balisage et de document balisé, nous renvoyons le lecteur à notre article (Merten 2005).

on fera appel à la *Text Encoding Initiative* (TEI) et à la *Corpus Encoding Standard* (CES).

La TEI a proposé depuis plus de dix ans des « lignes directrices » pour le codage et la représentation des documents dans les sciences humaines, pour la prose, la poésie, le théâtre, la transcription de *corpus* oraux, les dictionnaires... Le CES se fonde sur la TEI pour proposer des normes de balisage des différents types de *corpus*⁷. Et enfin, nous avons fait allusion au format TMX pour représenter les textes alignés.

La figure ci-dessous représente un passage d'un roman codé selon la TEI.



Figure 2 : The Moonstone, codé selon la TEI

Au départ, le document, diffusé par le projet Gutenberg, est au format texte. Une procédure de conversion permet de le mettre au format de la TEI. La procédure doit identifier au minimum les phrases et les paragraphes.

Prenons le document texte correspondant :

I am indebted to my dear parents (both now in heaven)
for having had habits¶

⁷ Pratiquement, ces initiatives ont développé des DTD, exploitables dans un éditeur approprié pour la création de tels documents. L'éditeur XML Oxygen, par exemple, permet de créer, entre autres, des documents TEI. On peut aussi se créer sa DTD personnalisée, en suivant le modèle « pizza » : <http://www.tei-c.org/pizza.html>.

of order and regularity instilled into me at a very early age. ¶

In that happy bygone time, I was taught to keep my hair tidy¶

at all hours of the day and night, and to fold up every article¶

of my clothing carefully, in the same order, on the same chair, ¶

in the same place at the foot of the bed, before retiring¶

to rest. An entry of the day's events in my little diary¶

invariably preceded the folding up. The "Evening Hymn"¶ (repeated in bed) invariably followed the folding up. ¶

And the sweet sleep of childhood invariably followed the¶

"Evening Hymn." ¶

On obtiendra une partie du codage simplement par une série de recherches et remplacements dans un traitement de texte ou un éditeur de texte :

deux marques de paragraphe sont remplacées par </s> </p> <p> <s>

la suite constituée d'un point et de deux espaces blancs est remplacée par </s> <s>

les marques de paragraphes restantes sont remplacées par un espace blanc.

Le balisage TEI code les éléments structuraux du document : les divisions en parties, paragraphes, phrases, les titres, etc. Il peut être nécessaire d'adjoindre aux mots du texte des annotations morphosyntaxiques, syntaxiques, voire sémantiques.

Le CES propose différents modes d'annotation des *corpus*. Un premier mode explicite le lien entre l'item et son annotation et en faisant deux sous-éléments du même élément :

```
<tok><orth>la</orth><ctag>DTN:sg</ctag></tok>
<tok><orth>mobilité</orth><ctag>SBC:sg</ctag></tok>
<tok><orth>linéaire</orth><ctag>ADJ:sg</ctag></tok>
<tok><orth>infinie</orth><ctag>ADJ:sg</ctag></tok>
<tok><orth>des</orth><ctag>DTC:pl</ctag></tok>
<tok><orth>& oelig ;
uvres</orth><ctag>SBC:pl</ctag></tok>
```

```
<tok><orth>de</orth><ctag>PREP</ctag></tok>
<tok><orth>Horta</orth><ctag>SBP:sg</ctag></tok>
```

Cette proposition n'est pas contraignante, et on peut préférer coder les informations linguistiques sous forme d'attributs. On aura un lien implicite entre l'item et son annotation :

```
<w ctag="DTN:sg">la</w>
<w ctag="SBC:sg">mobilité</w>
<w ctag="ADJ:sg">linéaire</w>
<w ctag="ADJ:sg">infinie</w>
<w ctag="DTC:pl">des</w>
<w ctag="SBC:pl">œuvres</w>
<w ctag="PREP">de</w>
<w ctag="SBP:sg">Horta</w>.
```

De la même manière, le CES suggère de distinguer le *corpus* et ses annotations, celles-ci renvoyant aux items concernés par des hyperliens. C'est ce qui est préconisé également pour l'établissement des équivalences linguistiques entre segments parallèles.

Les équivalences suivantes, représentées ci-après sous forme tabulaire :

C'est une vieille leçon de l'histoire :	<i>È un'antica lezione della storia :</i>
l'Empire impose sa loi au vaincu.	<i>l'Impero impone la sua legge ai vinti.</i>

devraient donc l'être comme suit selon le CES :

```
<s id="fr21">C'est une vieille leçon de l'histoire :</s>
<s id="fr22">l'Empire impose sa loi au vaincu.</s>
```

...

```
<s id="it01">È un'antica lezione della storia :</s>
<s id="it02">l'Impero impone la sua legge ai vinti.</s>
```

```
<translations>
<translation trans.loc="md-fr0505.xml"lang="fr"wsd="ISO8859-1" n="1" >>/>
<translation trans.loc="md-it0505.xml"lang="fr"wsd="ISO8859-1" n="2" >>/>
</translations>
```

```
<linkGrp targType="s">
<link>
  <align xlink:href="#fr21"/>
```

```

    <align xlink:href="#it01"/>
</link>
<link>
    <align xlink:href="#fr22"/>
    <align xlink:href="#it02"/>
</link>
</linkGrp>

```

Le format TMX emploie un lien implicite : les segments équivalents font partie de la même unité de traduction :

```

<tu tuid="26" segtype="sentence">
<tuv lang="FR-FR"><seg>C'est une vieille leçon de
l'histoire :</seg>
</tuv>
<tuv lang="IT-IT"><seg>È un'antica lezione della
storia :</seg>
</tuv></tu>

```

```

<tu tuid="27" segtype="sentence">
<tuv lang="FR-FR"><seg>l'Empire impose sa loi au
vaincu.</seg>
</tuv>
<tuv lang="IT-IT"><seg>l'Impero impone la sua legge ai
vinti.</seg>
</tuv></tu>

```

En associant une feuille de style, on peut l'afficher sous forme tabulaire, ce qui facilitera considérablement la consultation et les recherches :

Corpus "Monde diplomatique: Ramonet mai 2005" : 31 exemples

Néo-impérialisme	Neo imperialismo
Le plus stupéfiant n'est pas tant l'indécence du propos que la manière résignée, apathique, dont les grands médias ont couvert l'installation de celui qu'il faut bien appeler le « proconsul des Etats-Unis ».	Stupefacente, più ancora dell'indecenza della frase, l'apatica rassegnazione dei grandi media nell'annunciare l'insediamento di quello che dobbiamo pur chiamare il « proconsole degli Stati Uniti ».
Comme s'il n'y avait plus de droit international.	Come se il diritto internazionale non esistesse più.
Comme si nous étions revenus à l'époque des mandats (1).	Come se fossimo tornati all'epoca dei mandati .

Figure 3 : fichier TMX avec feuille de style XSL

La feuille de style est reprise en annexe.

6. L'EXTRACTION DE CHAÎNES

Toutes ces procédures ne seront d'aucune utilité si on ne peut extraire des *corpus* les chaînes intéressantes. Le problème est qu'il faut pour cela accéder à des informations linguistiques. On a vu comment les représenter, encore faut-il les établir.

6.1. La catégorisation

On fait pour cela appel à un catégoriseur ou étiqueteur, qui déterminera la catégorie morphosyntaxique de chaque mot. Nous avons utilisé WinBrill⁸. Le résultat de l'étiquetage se présente comme suit :

```
la/DTN:sg mobilité/SBC:sg linéaire/ADJ:sg infinie/ADJ:sg
des/DTC:pl œuvres/SBC:pl de/PREP Horta/SBP:sg
```

Ce résultat a été converti au format de la CES, pour se conformer aux standards et faciliter les traitements ultérieurs, mais cette opération n'était pas indispensable. La conversion ou le repérage des chaînes intéressantes, par exemple une succession de deux adjectifs, n'est pas déterministe ici, mais contextuelle. On ne pourra se contenter de la fonctionnalité « recherche et remplacement ». Il faut utiliser un langage de « pattern matching », comme *Perl*.

6.2. Extraction de chaînes et expressions régulières

Nous avons extrait des chaînes intéressantes du fichier au format CES exposé plus haut au moyen de la commande Grep de Linux. La commande

```
grep "<w ctag =\"ADJ:[sp] [gl]\"[[ :alnum:]]+</w> <w
ctag=\"ADJ\""
```

repère les suites de deux adjectifs. Cela pourrait se faire aussi avec une commande écrite en *Perl* ou tout autre langage permettant de manipuler des expressions régulières.

Frantext, cité plus haut, propose une version catégorisée permettant de rechercher n'importe quel mot d'une certaine catégorie, même s'il ne s'agit pas vraiment d'expressions régulières :

```
&e (g=S) &e (g=A) &e (g=A)
```

permet d'extraire les séquences substantif-adjectif-adjectif. Mais il n'y aurait pas moyen de rechercher une suite quelconque d'adjectifs. Il faudra donc rechercher successivement les suites de deux, de trois, voire de quatre adjectifs.

6.3. Les concordanciers

Un « simple » concordancier peut déjà être très utile en listant les mots en contexte. Mais le repérage des exemples intéressants devra se faire manuellement.

⁸ *WinBrill* est disponible sur le site de l'Inalfr : <http://www.inalfr.cnrs.fr/>.

ous les éléments constitutifs . pour des multiples d ' un même atome , on multiplie les liaisons ? et qui contiennent des liaisons multiples seront discutés dans les chapitres side dans leurs structures : ce sont des multiples d ' une unité à cinq atomes de carbone entent souvent des maxima d ' absorption multiples (figure 11 . 14) . Problème 11 . pports d ' intégration exacts . Les pics multiples sont approximativement symétriques aucun risque de réactions d ' alkylation multiples (voir section 12 . 4 . 1) . Par exemple , amino , alcoyle , et les liaisons multiples carbone - carbone , sont repérés par d ' autres substituants - les liaisons multiples carbone - carbone ainsi que les groupes pharmaceutiques . 17 . 3 Stéréocentres multiples Jusqu ' ici , la discussion des composés deux unités d ' ATP : EQUATION (étapes multiples) Acétyl - CoA Le processus global

Figure 4 : concordance obtenue dans Lexico 3

6.4. Les corpus d'exemples : XML ou Access ?

Une fois les exemples intéressants collationnés, on est en droit de se demander s'il est préférable de les stocker dans un document XML ou dans une base de données. XML a la réputation d'être à la frontière entre le document et la base de données. Certes nous aurions tendance à plaider en faveur de formats standards, mais il faut bien reconnaître que la manipulation de tels fichiers n'est pas toujours évidente et qu'on ne peut y multiplier les voies d'entrée. Ainsi, on voudra, toujours dans notre contexte d'étude, lister les exemples par structure et par adjectif, pour repérer quels sont les adjectifs que l'on trouvera en antéposition et en postposition. On peut écrire une feuille de style XSL qui trie les exemples sur l'un ou l'autre attribut, clé ou élément. Mais il n'est pas possible d'avoir un double tri. De ce point de vue, une base de données relationnelle comme Access est beaucoup plus souple, à condition d'adopter le schéma de base de données ci-après :

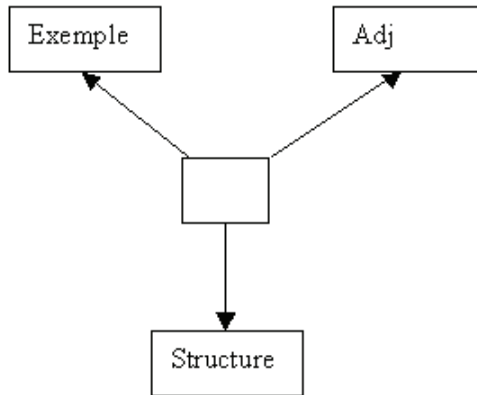


Figure 5 : schéma de base de données

5. LES RÈGLES ET L'AUTOMATISATION

Le langage XSLT est remarquablement puissant, même si on peut regretter sa verbosité ainsi que l'absence de certaines structures de données, comme les tables associatives. Il nous a permis, en utilisant un analyseur (Saxon en l'occurrence), de

tester nos hypothèses de travail en réalisant un véritable petit programme de traduction automatique.

Soit l'arbre correspondant au groupe néerlandais « Internationaal gerechtshof » :

```
<sn id_sn="nl_sn2" langue="nl" nombre="sg" genre="n"
id_tete="nl_hof">
<car cat="adj" valeur1="class" valeur2="origine"
valeur3="generique" source="nl_internationaal" pos-
absolue="ante" pos-relative="2"> internationaal </car>
<car cat="gen" valeur1="class" valeur2="finalite"
source="nl_gerechts" pos-absolue="ante" pos-
relative="1"> gerechts </car>
<tete cat="nom" source="nl_hof"> hof </tete>
</sn>
```

Le correspondant en français sera :

```
<sn langue="fr" genre="f" nombre="sg" id_tete="fr_cour"
regle="r1" sem_car1="class_origine_generique">
<tete cat="nom" source="fr_cour" valeur="coll"
nombre="sg" genre="f">cour</tete>
<car source="fr_international" valeur1="class"
valeur2="origine" valeur3="generique" pos_r="1"
pos_a="post">internationale</car>
<car source="fr_de_justice" valeur1="class"
valeur2="finalite" valeur3="" pos_r="2" pos_a="post">de
justice</car>
</sn>
```

XML est particulièrement approprié à la représentation d'arbre et d'attributs. Moyennant les dictionnaires (en XML) et les procédures (en XSLT) appropriés, on peut évaluer la pertinence des règles et des informations grammaticales nécessaires dans la compréhension d'un problème linguistique.

6. CONCLUSION

Il n'est pas toujours aisé de rassembler les exemples intéressants pour l'étude d'un phénomène linguistique, surtout si cette étude se veut, en plus, comparative. Divers moyens existent pour exploiter des *corpus* existants et pour se constituer ses propres *corpus*. Ces moyens ne sont pas toujours simples, et peuvent exiger une compétence informatique que n'ont pas toujours les linguistes. On ne peut que plaider pour une formation à ces outils, tout autant que pour le développement et la diffusion d'outils standards. Nous espérons, en témoignant de notre expérience et des techniques utilisées, en transcrivant le code utilisé pour certaines applications, apporter notre petite pierre à cet édifice.

BIBLIOGRAPHIE

- Bray (T.), Paoli, (J.), Sperberg-McQueen, (C. M.), Maler, (E.) édit., 2004 : *Extensible Markup Language (XML) 1.0 (third edition) W3C Recommendation 04 February 2004*, <http://www.w3.org/TR/2004/REC-xml-20040204/>.
- Cinque (G.), 1994 : « On the Evidence for Partial N-Movement in the Romance DP », dans Cinque (G.), Koster (J.), Pollock (J.-Y.), Rizzi (L.) et Zanuttini (R.) édit., *Paths Towards Universal Grammar. Studies in Honor of Richard S. Kayne*, Washington, D.C., Georgetown University Press.
- McEnery (T.), Wilson (A.), 1996 : *Corpus Linguistics*, « Edinburgh, Edinburgh Textbooks in Empirical Linguistics », Edinburgh University Press,
- Merten (P.), 2005 : « La traduction de documents XML dans les mémoires de traduction », dans *Équivalences*, 32-1, Bruxelles.
- Rijkhoff (J.), 1988 : « Order in the Noun Phrase of the Languages of Europe », dans Siewierska (A.) édit., *Constituent order in the Languages of Europe*, Berlin-New York, Mouton de Gruyter, pp. 321-382.
- Salkoff (M.), 1999 : *A French-English Grammar. A Contrastive Grammar on Translational Principles*, Amsterdam-Philadelphia, John Benjamins Publishing Company.
- Siewierska (A.), 1988 : *Word Order Rules*, London-Sydney-New York, Croom Helm.
- Sperberg-McQueen (C. M.) and Burnard (L.) édit., 1994-2002 : *Guidelines for Electronic Text Encoding and Interchange (TEI P5)*, TEI P5, Oxford-Providence-Charlottesville-Bergen, <http://www.tei-c.org/P5/>.
- Teysier (J.), 1968 : « Notes on the syntax of the adjective », dans *Lingua* 20, pp. 225-249, Amsterdam, North-Holland Publishing Co.
- Wilmet (M.), 2003³ [1997] : *Grammaire critique du français*, Bruxelles, Duculot.
- XCES, 2002 : *Corpus Encoding Standard for XML*, Vassar College-Loria/CNRS, Poughkeepsie NY – Vandœuvre-lès-Nancy, <http://www.cs.vassar.edu/XCES/>.

ANNEXE : FEUILLE DE STYLE XSL

```

< ? xml version = "1.0" encoding = "UTF-8"? >
<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:template match="/">
    <html><head><style type="text/css">
        body {}
        h4 {font-family:Verdana; font-size:12pt; color:darkred; text-align:left}
        p {font-family:Verdana; font-size:10pt}
    </style></head>
<body><table border="1">
<xsl:for-each select="tmx">
    <h4>Corpus "<xsl:value-of select="/tmx/header/@corpus_name" />"
    <xsl:value-of select="header/@corpus_langues" /> : <xsl:value-of
select="count(body/tu)" /> exemples</h4>
<xsl:apply-templates />
</xsl:for-each></table></body></html>
</xsl:template>
<xsl:template match="tu">
    <tr>
    <xsl:apply-templates />
    </tr>
</xsl:template>
<xsl:template match="tuv">
    <td><p>
    <xsl:apply-templates />
    </p></td>
</xsl:template>
<xsl:template match="source">
    <tr>
    <xsl:apply-templates />
    </tr>
</xsl:template>
</xsl:stylesheet>

```

INTERPRÉTARIAT À DISTANCE, COLLECTE ET DIFFUSION COLLABORATIVES DE *CORPUS* ORAUX BILINGUES, EN SITUATION OUVERTE, FINALISÉE OU CONTEXTUALISÉE

Georges Fafiotte
GETA, laboratoire CLIPS-IMAG
Grenoble, France
georges.fafiotte@imag.fr

Résumé : Sont présentés ici les motivations et les résultats actuels de l'action de développement d'un Environnement réseau pour l'interprétariat multimodal : les plateformes Erim-Interprète, pour la communication orale multilingue sur réseau, et ERIM-Collecte pour la collecte de dialogues oraux spontanés bilingues traduits. Quelque 35 heures de corpus ont déjà été collectées, en français-chinois, français-vietnamien, français-tamoul et français-hindi. Nous relatons les conditions et difficultés particulières de ces collectes. Il est proposé de mettre ces ressources en ligne, pour que se crée une dynamique de production collaborative, de mutualisation et d'enrichissement de ressources linguistiques.

Mots-clés : interprétariat à distance sur réseau, dialogues bilingues spontanés, corpus oraux bilingues traduits, collecte collaborative, mutualisation de ressources linguistiques

1. INTRODUCTION

À l'oral comme précédemment à l'écrit, le multilinguisme est un enjeu central des interactions à distance, en particulier pour le développement d'environnements d'aide à la communication multilingue sur réseau, pour la construction de bases lexicales multilingues, généralistes ou de spécialité, pour les recherches en traduction automatique, et bien sûr en terminologie.

De plus, deux questions sensibles, sur ces domaines d'investigation, sont celles d'un équilibre à construire entre les grandes langues véhiculaires, parmi lesquelles le français, et celle également d'un maintien de la diversité des langues d'origine des partenaires de conversations (« conversants »), par exemple grâce à des services peu coûteux d'interprétariat intermittent à distance sur Internet.

Nous développons une famille de plateformes logicielles ERIM (Environnement sur réseau pour l'interprétariat multimodal), permettant entre autres d'une part la communication multilingue sur réseau, traduite à distance par interprète, d'autre part la collecte et la réécoute de dialogues spontanés oraux bilingues ; ces plateformes ont été déjà utilisées en français-chinois, français-vietnamien, français-tamoul et français-hindi.

Sont présentées d'abord ici les motivations premières de cette action, puis l'environnement ERIM et ses facettes actuelles. Nous rendons compte ensuite des premières collectes expérimentales réalisées avec ERIM en quatre couples de langues « français – langue L », et des conditions particulières ou difficultés rencontrées lors de telles collectes. Il est proposé enfin de mettre ces ressources en ligne, pour contribuer à créer une dynamique de production collaborative, de partage et d'enrichissement de ressources linguistiques.

1. PREMIÈRE MOTIVATION, LE CONSTAT D'UN MANQUE DE CORPUS DE PAROLE SPONTANÉE

Nous constatons le manque notoire de grands *corpus* de dialogues oraux spontanés bilingues, traduits, en français et différentes langues, en langue de spécialité comme en langue généraliste, disponibles en accès libre de droits. De telles ressources sont utiles à diverses recherches sur le patrimoine linguistique, essentielles également au développement de systèmes de traduction automatique de la parole (TAP).

L'importance de grands *corpus* réalistes est essentielle en particulier pour la construction de systèmes de TAP. Il est besoin, pour établir des modèles de langues parlées en situations réelles (modèles très différents de ceux des dialogues monolingues et de la langue écrite), de *corpus* parallèles d'énoncés transcrits alignés. Très peu de ressources de ce type sont produites (NEC, ATR, ELRA, quelques autres), et encore moins sont disponibles en accès libre.

Pourquoi ces *corpus* sont-ils « propriétaires » ? Parce qu'ils sont coûteux à collecter, et encore plus à transcrire et annoter. Les rendre accessibles gratuitement semble déraisonnable s'ils ont été beaucoup « travaillés ».

Nous voulons donc pouvoir collecter, à coût réduit, de grands *corpus* de parole traduite « français-langue L », et les proposer en accès libre aux chercheurs intéressés, sur le Web (Fafiotte 2004). Le projet AUF-LTT VTH-Fra. Dial se propose actuellement de valider cette hypothèse.

Ces *corpus* seront « bruts » (fichiers de parole et de métadonnées, descripteurs des conditions de collecte), sans transcription ni annotations. L'enrichissement des *corpus*, leur utilisation pour des recherches linguistiques, ou pour créer des « modèles de langue (s) parlée (s) », pourront être réalisés ensuite librement par des membres de la communauté scientifique.

Nous proposons, pour favoriser ce processus de création et de partage de telles ressources linguistiques, d'offrir à des interprètes en perfectionnement, et à des locuteurs, l'accès à une plateforme logicielle sur Internet, et un service de traduction de bonne qualité et peu coûteux, « en l'échange » d'une collecte de *corpus* de dialogues, dans des situations ou des contextes linguistiques facilement ajustables par des chercheurs : situations ouvertes et langue généraliste, domaines finalisés et langues de spécialité, situations contextualisées.

Nous souhaitons mettre à disposition, en libre accès sur Internet, les *corpus* que nous collectons, et certains des logiciels développés (à commencer par la plateforme de col-

lecte et le composant de réécoute sélective de *corpus*), pour favoriser un volontariat de production de ressources linguistiques (*corpus* bilingues de parole spontanée, en toutes langues) partageables au sein de différentes communautés de chercheurs.

2. DÉVELOPPER LA COMMUNICATION ORALE MULTILINGUE SUR RÉSEAU

Notre seconde motivation centrale est le développement de l'expression orale multilingue sur réseau, entre locuteurs monolingues ou partiellement bilingues, médiatisée par des interprètes humains dans un premier temps, puis utilisant progressivement des aides automatiques appropriées, héritées du traitement automatique des langues naturelles (TALN) ou de la traduction assistée de la parole (TAP) — aides aux locuteurs et aides aux interprètes.

Il peut s'agir d'un interprétariat en situation d'urgence (tel que le pratique par téléphone en France l'association STU (« Service de traducteurs d'urgence »), 24 heures sur 24 et 7 jours sur 7, en 85 langues), d'un service d'assistance linguistique à l'expression orale multilingue (*hot line* sur domaines finalisés, médical, juridique...) ou encore d'un interprétariat intermittent sur rendez-vous.

Il est souhaitable que se développent de tels services d'interprétariat à faible coût sur Internet, et qu'ils favorisent les échanges entre locuteurs « peu bilingues » ou monolingues, et le maintien de la diversité des langues d'origine, en particulier la pratique des langues peu dotées. Il est possible d'espérer qu'ils constituent un relatif facteur d'équilibrage entre grandes langues véhiculaires, parmi lesquelles le français.

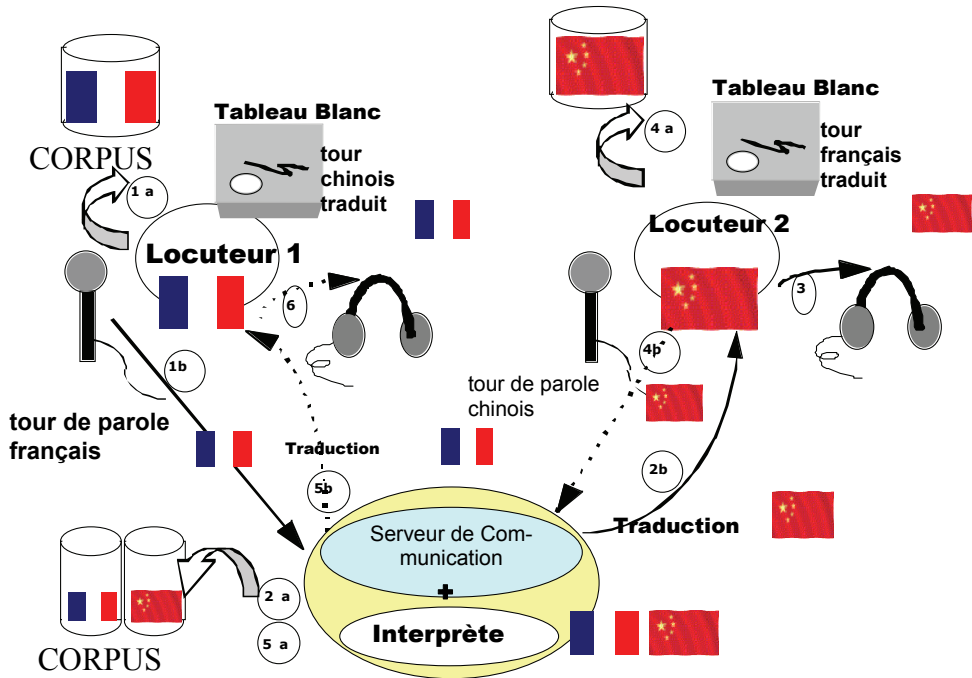
Après collecte puis transcription de dialogues spontanés, dans les langues orales utilisées en situations réelles (transcriptions manuelles dans un premier temps, à terme post-éditées à partir de prétranscriptions effectuées par reconnaissance vocale), tout traitement linguistique est possible, par exemple l'extraction de bases terminologiques multilingues, spécialisées ou non, en différents contextes de langues parlées, de dialectes, de parlers régionaux.

3. ERIM, UN ENVIRONNEMENT POUR L'INTERPRÉTARIAT MULTIMODAL SUR INTERNET, AVEC COLLECTE POSSIBLE DE DIALOGUES PARLÉS BILINGUES TRADUITS

Dans le cadre du projet ERIM (Environnement réseau pour l'interprétariat multimodal), a été conçue et prototypée une famille de plates-formes pour la communication multilingue à distance. Deux sont actuellement opérationnelles, et en validation d'utilisabilité et de robustesse : ERIM-Interprète, pour l'interprétariat multimodal sur le Web (et « socle » des autres composants), et ERIM-Collecte pour la collecte de *corpus* de dialogues parlés traduits, qui permet d'enregistrer, de manière anonyme et avec l'accord des intervenants, toutes les interactions de dialogues spontanés bilingues.

ERIM-Collecte, en configuration avec ℓ locuteurs et i interprètes (ici $\ell = 2, i = 1$)

ERIM-Interprète permet d'explorer des scénarios assez différents de ceux de l'interprétariat classique :



- « interprétariat intermittent à la demande » : les locuteurs essaient de converser en utilisant la connaissance qu'ils ont de la langue de leur interlocuteur, ou dans une langue véhiculaire, ou connue des deux ; lorsque cette communication s'avère impraticable, ou pour des séquences « sensibles » de leur échange, ils font appel momentanément aux services d'un interprète disponible sur le Web, qui peut les aider ;
- téléconférence (*conference call*) : les interlocuteurs prennent un rendez-vous avec un interprète pour un créneau de durée donnée.

D'autre part, nous avons développé ERIM-Collecte pour

- collecter à coût réduit, à partir de dialogues réels, des données « brutes » que d'autres équipes traiteront ensuite, pour construire par exemple des bases lexicographiques ou terminologiques, des modèles de langue orale pour la TA de parole ;
- proposer à des volontaires de produire ces données gratuitement... en l'échange d'un accès libre aux plates-formes ERIM nécessaires et aux services qu'elles proposent.

Il est possible d'utiliser ces environnements pour créer des situations variées de saisie de l'expression orale, en langue de spécialité, en domaine ciblé, dans des

contextes linguistiques particuliers (par exemple variantes lexicales entre parlers vietnamiens du nord ou du sud).

En multimodalité, sont actuellement possibles l'échange de textes courts, le partage sur un « tableau blanc » de documents textuels et graphiques, avec pointages, marquages et soulignements libres (Loken-Kim *et al.* 1994), et la communication vidéo.

Une ressource de réécoute sélective, le module « AuditDial », permet de reconstituer tout ou partie du dialogue, chronologiquement ou avec extraction de versions monolingues. Il facilite un suivi visuel de l'échange. Il fonctionne également en version bimodale (tours de parole, échange de textes courts), et une extension est en cours au tableau blanc, pour des *corpus* comprenant la collecte des événements de multimodalité. Il sera accessible sur un site Web (DistribDial) d'accès aux *corpus* produits, pour des utilisateurs de la communauté scientifique.

En revanche ERIM-Collecte, destinée par choix initial à la constitution de *corpus* de données brutes (fichiers descripteurs de session et tours de parole) ne propose pas de ressource particulière d'aide intégrée à la transcription, ni à l'annotation. Ces dernières sont réalisables hors ERIM.

Une variante ERIM-TA, prototype orienté vers un service de traduction partiellement automatique de parole aidée par le locuteur (et en même temps plateforme « banc d'essai de composants » pour l'ingénierie de TAP) est en cours de développement avancé.

La facette ERIM-Aides de l'environnement, orientée vers des aides en ligne à la communication multilingue sur réseau, est en début de prototypage. Les aides aux « interprètes en ligne » (professionnels ou occasionnels) peuvent être lexicales, terminologiques, de prononciation... Il peut s'agir également d'aides à l'interaction (multimodalité, utilisation d'un tableau blanc pour le partage entre locuteurs de documents visuels ou graphiques avec marquages à main levée, vision de l'interlocuteur par *webcam* en situation de dialogue à deux ou pour des applications de visioconférence multilingue, transcription instantanée pour rappel visuel du texte des derniers tours de parole...), ou d'aides à la gestion de l'activité de l'interprète, à la prise et à la consultation (sur le vif, ou après et avant une séquence de traduction orale) de notes personnelles, écrites ou orales...

Les aides aux locuteurs sont comparables, avec à terme l'adjonction de composants de « traduction partiellement automatique » de parole (traduction automatique « aidée par l'humain » (Boitet, Blanchon 1994)) facilitant une communication médiatisée (de la meilleure qualité possible) dans des situations où il faudrait passagèrement se débrouiller sans l'interprète.

L'architecture ouverte de la plateforme générique — nombre quelconque de locuteurs, d'interprètes, possibilité d'intégration (*plug-in*), de composants externes — facilite des développements futurs. Nous prévoyons d'utiliser par exemple les connaissances vocales intégrées d'ERIM-TA pour produire ces prétranscriptions en

version instantanée brute, à la mesure certes de la qualité des composants de reconnaissance disponibles.

Erim-Formation, variante pour l'*e*-training à distance en interprétariat bilingue, pour des étudiants interprètes ou des interprètes en stage de perfectionnement, est actuellement en test de faisabilité sur la plate-forme générique de base.

4. COLLECTE DE *CORPUS* BILINGUES : PROJETS LIAMA-CHINFADIAL, AUF-LTT VTH-FRA. DIAL

4.1. Premières actions de collecte de dialogues oraux bilingues finalisés

ERIM-Collecte a été utilisé dans le cadre du projet Liama ChinFaDial, en partenariat avec le NLPR (National Laboratory for Pattern Recognition — Institut d'automatique de l'Académie des sciences de Chine, à Pékin), pour la collecte de dialogues parlés bilingues français-chinois, spontanés et traduits, entre deux locuteurs monolingues (Fafiotte *et al.* (2004)). Les situations sont finalisées (réservation hôtelière), en dialogue spontané. Une dizaine d'heures de dialogues oraux ont été enregistrées lors de sessions au CLIPS et au NLPR, et constituent un premier *corpus* de données brutes, non transcrites ; les collectes se poursuivent.

Dans le cadre du projet VTH-Fra. Dial soutenu par l'AUF (réseau LTT), d'autres *corpus* sont actuellement en cours de collecte dans des régions francophones des pays concernés (une vingtaine d'heures à ce jour), en français-vietnamien (au CLIPS et aux laboratoires DATIC de l'Université de Danang et MICA de Hanoï), français-hindi (au laboratoire CFILT, « Center for Indian Languages Technology », de l'IIT-Bombay) et français-tamoul (au CLIPS et à Pondichéry).

Des actions complémentaires pourraient concerner la collecte contrastive de parlers régionaux et de la langue principale ou nationale, avec éventuelle médiation d'un interprète.

4.2. Conditions et vicissitudes de la collecte de terrain de dialogues spontanés bilingues

Au vu de notre expérience de collecte, qui porte actuellement sur trente à trente-cinq heures de dialogues oraux spontanés traduits par interprète, en quatre couples de langues (français–langue L), nous proposons quelques idées pour contribuer au développement d'une dynamique de collecte collaborative à coût réduit, et de mutualisation des ressources linguistiques produites.

Même si l'organisation de collectes avec des participants bénévoles n'est pas une tâche aisée, nous nous efforçons dans une telle optique, de cerner les conditions les plus favorables de cette activité — qu'elles concernent les paramètres logistiques, l'organisation même, la disponibilité des intervenants, leur motivation, les « attitudes » incitatives.

Logistique

Selon les situations, nous utilisons des configurations matérielles différentes, pour les collectes classiques à 2 locuteurs et 1 interprète. À distance, nous utilisons volontiers la configuration idéale, à 3 stations séparées (sur réseau local le plus souvent, ou sur le Web), chaque intervenant étant donc isolé phoniquement de ses partenaires de dialogue.

Une configuration particulière à 2 stations séparées (logistique plus simple) est dite « de situation de visite », qui fait partager la même station aux 2 locuteurs (l'un rendant visite à l'autre) : ils se passent le micro et écoutent les haut-parleurs incorporés (ou dialoguent avec 2 microcasques reliés aux mêmes connecteurs). Concrètement, l'absence parfois de réseau local efficace, ou de lieux distincts, conduit à installer cette « situation de visite » dans une pièce unique, en connectant les deux ordinateurs par câble ethernet croisé.

Enfin, une logistique réduite, adaptée à des conditions de terrain plus sommaires et finalement assez fréquentes, conduit à travailler à 3 intervenants sur une même station (2 locuteurs monolingues et l'interprète), souvent sur ordinateur portable, chacun prenant à son tour le micro ou parlant dans un « micro de table ronde » triple : on peut ainsi ne pas émettre les tours de parole sur haut-parleurs, puisque les interlocuteurs interviennent « de vive voix » et s'entendent, ce qui permet un rythme de collecte plus soutenu, et une conversation plus naturelle.

Organisation

Il s'agit de planifier des campagnes de collecte, sur la base la plus régulière possible, avec les difficultés inhérentes aux situations de bénévolat.

Disponibilité

Le locuteur en langue maternelle du lieu où se tient le dialogue est en général relativement aisé à trouver, sinon à mobiliser — et le locuteur dans l'autre langue (en particulier le français, lors de collectes à l'étranger) souvent beaucoup moins.

Motivations, incitations, selon le profil des intervenants

Qui sont les interprètes et les locuteurs, et quelles sont leurs motivations pour une action bénévole ? Leur participation est-elle volontaire, simplement curieuse, parfois relativement « contrainte » ? Quels aménagements ou quels dispositifs peuvent les inciter à participer ?

<i>Profil d'intervenants</i>	Participation (en nombre)	Motivation	Dispositif souhaitable
chercheurs impliqués dans cette action, chercheurs du domaine	+	intérêt personnel lié à leur recherche, curiosité,	une bonne planification
autres chercheurs, ingénieurs de recherche	-	curiosité, participation « solidaire »	une bonne information sur la recherche en cours, une bonne communication, une très bonne planification
administratifs	-	gratification « qualitative » ? obligation ?	incitation institutionnelle (responsable hiérarchique)
étudiants, stagiaires	+++	peu « spontanée », rarement volontaire, gratification « académique » ?	stages, incitation institutionnelle (« crédits » dans leur cursus, pour participation active)
enseignants (de français) à la retraite	variable selon les contextes, parfois ++	grande disponibilité, intérêt « spontané », gratification « symbolique » (participation à une recherche)	une bonne information sur la recherche en cours, une bonne communication, effectuer des auditions
interprètes professionnels bénévoles, en perfectionnement	-	situation de perfectionnement volontaire ?	proposer l'accès <i>a posteriori</i> au <i>corpus</i> enregistré, une très bonne organisation

Tableau 1 : Intervenants, leurs motivations, les incitations possibles, en collecte de *corpus* oraux bilingues

Il s'avère souvent bien difficile de trouver des participants à la fois volontaires et bénévoles (Bellinck *et al.* 2005). Le milieu des étudiants doctorants ou en Master offre une combinaison de facteurs motivationnels intéressants : bonne acceptation du bénévolat, curiosité, expérience perçue comme enrichissante ou formatrice, avec s'il est possible une gratification « académique » lorsque des responsables de leurs filières l'acceptent. Il sera donc efficace, dans une stratégie de collecte participative, de développer des dispositifs incitatifs en milieu étudiant (étudiants avancés en langues, en interprétariat, doctorants étrangers...), et éventuellement auprès

d'interprètes en demande d'activités de perfectionnement et de *training*, de même que de solliciter d'anciens enseignants de langues.

5. PARTAGE ET UTILISATION DES RESSOURCES LINGUISTIQUES PRODUITES

Nous préparons actuellement la mise à disposition, en accès libre avec licence de type GPL, sur 2 sites Web avec serveur basé au GETA-CLIPS,

- d'une part, des *corpus* déjà collectés sur un site DistribDial, après validation des conditions de collecte, des métadonnées et des données, et
- d'autre part, sur un site CollectDial, de la plate-forme de collecte ERIM-Collecte et du composant de réécoute sélective AuditDial, après validation (en cours) de leur utilisabilité, des interfaces-utilisateur, et de leur robustesse.

Des tests d'utilisation distante, à moyenne distance (100 km), ont été effectués, et d'autres tests vont intervenir sur très longue distance, qui doivent repérer les conditions techniques et logistiques minimales d'une bonne communication et d'une bonne collecte (débit des liaisons sur réseau, créneaux horaires...).

Nous souhaitons ainsi contribuer à promouvoir des actions de mutualisation de ressources : *corpus* parallèles de dialogues parlés spontanés, à collecter, puis à transcrire et annoter, en participation collaborative.

Les chercheurs en ingénierie de la TA de parole, en lexicographie et lexicologie, en terminologie (notamment en situation de langue de spécialité, ou de contextualisation linguistique), peuvent spécifier des scénarios ou des protocoles de capture appropriés, et surtout travailler en aval, à partir des ressources linguistiques produites et partagées.

La variante ERIM-Formation prévoit de proposer à des étudiants avancés en interprétariat bilingue, ou à des interprètes en perfectionnement, différents modes de formation à distance (FAD) sur le Web, de type *e-training*, pour une activité en ligne « en direct » ou « de doublage ». Toute intervention est enregistrable (avec l'accord des intervenants) et consultable, et un travail pédagogique intéressant peut s'ensuivre.

Cette ressource rejoint l'approche collaborative du projet ERIM, car il est attendu de la part d'interprètes de ce profil qu'ils acceptent de se porter volontaires, pour coopérer à l'activité collectrice sur ce type d'outil « d'apprentissage à distance par la pratique », contribuant à la création peu coûteuse de *corpus* de dialogues spontanés, rendus anonymes.

6. CONCLUSION

Deux plates-formes ERIM sont actuellement opérationnelles (Erim-Interprète et ERIM-Collecte), utilisées respectivement pour la communication multilingue sur réseau médiatisée par interprète humain, et pour la collecte de dialogues oraux

spontanés bilingues, une trentaine d'heures à ce jour, en français-chinois, français - vietnamien, français-tamoul et français-hindi. Une évaluation de leur utilisabilité et de leur robustesse est en cours. Une extension de la multimodalité est attendue.

Deux sites en accès libre sous licence de type GPL, DistribDial et CollectDial, permettront à la communauté scientifique d'une part la consultation de *corpus* bruts collectés (tours de parole et métadonnées) ou un travail d'enrichissement, d'autre part la collecte de nouveaux dialogues bruts.

Il nous apparaît que de tels outils (environnements logiciels pour l'interprétariat, la collecte de dialogues oraux, l'aide à la communication multilingue, la formation), et leur mise à disposition en téléaccès libre, au-delà du développement synergique qu'ils permettent (Blanchon 2004) - en premier lieu un développement de services de communication multilingue multimodale à coût réduit sur le Web, notamment entre langues peu dotées et grandes langues véhiculaires -, peuvent sous-tendre un effort de production de ressources linguistiques (documents multimédias multilingues et multimodaux, *corpus* de dialogues spontanés bilingues parallèles, et par extraction lexicale, bases terminologiques spécialisées), favorisant une création collaborative et une mutualisation de ressources linguistiques.

Ces actions peuvent être conduites dans le contexte de recherches fondamentales, en linguistique, en ingénierie de la TAP, en vue de la préservation de patrimoines linguistiques, ou dans le contexte d'un appui à la Francophonie - par exemple en favorisant l'apprentissage de l'interprétariat bilingue à distance, en facilitant la découverte ou la pratique des langues sur le Web.

REMERCIEMENTS

Ces travaux ont été soutenus par le CLIPS-IMAG (Université Joseph Fourier — Grenoble I, INPG, CNRS), la région Rhône-Alpes (projet ERIM), le laboratoire franco-chinois LIAMA (projet ChinFaDial), et le réseau LTT de l'AUF (projet VTH-Fra. Dial).

Remerciements à Zhai JianShe (Université de Nankin, Chine) et Julien Lamboley (INSA-Lyon) qui ont contribué de manière centrale au développement des plateformes, ainsi qu'aux collègues du CLIPS et de MultiCom (à Grenoble), du National Laboratory for Pattern Recognition (CAS-IA à Pékin, Chine), de l'Université de Danang (Vietnam), de l'Alliance française de Pondichéry et de l'IIT-Bombay (Inde), qui ont participé aux expérimentations et collectes.

BIBLIOGRAPHIE

Bellynck (V.), Boitet (C.), Kenwright (J.), 2005 : « Construction collaborative de lexiques techniques et de contextes d'usage par des apprenants : un service Web utilisable en Francophonie », dans *Actes des 7^{es} Journées scientifiques AUF – LTT, Bruxelles*, 8-10 sept. 2005.

- Blanchon (H.), 2004 : « Comment définir, mesurer et améliorer la qualité, l'utilisabilité et l'utilité des systèmes de TAO de l'écrit et de l'oral — une bataille contre le bruit, l'ambiguïté, et le manque de contexte », dans *HDR UJF*, 20 déc. 2004, 320 pp.
- Boitet (C.), Blanchon (H.), 1994 : « Multilingual Dialogue-Based MT for Monolingual Authors : the LIDIA Project and a First Mockup », dans *Machine Translation vol. 9 (2)*, pp. 99-132.
- Fafiotte (G.) 2004: « Building and sharing multilingual speech resources, using ERIM generic platforms », dans *Proc. Coling-MLR 2004, Genève*, 28 Aug. 2004.
- Fafiotte (G.), Boitet (C.), Seligman (M.), Zong (C.-Q.), 2004: « Collecting and Sharing Spontaneous Speech Corpora: the ChinFaDial Experiment », dans *Proc. LREC 2004, Lisboa*, 26-28th May 2004.
- Loken-Kim (K.-H.), Yato (F.), Morimoto (T.), 1994 : « A Simulation Environment for Multimodal Interpreting Telecommunications », dans *Proc. IPSJ-AV workshop*, March 1994.
- Site web C-STAR : <http://www.c-star.org>
- Site web Nespole! : <http://nespole.itc.it>
- Site web Papillon : <http://www.papillon-dictionary.org>

INDEXATION HUMAINE ET INDEXATION AUTOMATISÉE : LA PLACE DU TERME ET DE SON ENVIRONNEMENT

Mustafa El Hadi Widad

UFR IDIST/CERSATES, UMR 8529

Université Lille III

Lille, France

Résumé : Dans cet article, nous examinerons l'apport de la terminologie en accès à l'information. Nous ferons tout d'abord un bref rappel de la problématique de l'indexation et de la recherche d'information et soulignerons ensuite l'apport de la terminologie dans ces deux processus interdépendants. Nous retracerons l'évolution des techniques en acquisition de ressources terminologiques et leur apport spécifique à l'indexation humaine et automatique ainsi qu'à la construction de langages documentaires tels que les thesaurus.

Mots-clés : indexation, extraction de termes et de relations sémantiques, thesaurus, accès à l'information

1. INTRODUCTION

Toute activité dans les sociétés industrialisées produit et consomme de l'information. Comme celle-ci se cristallise, en particulier en vue ou à l'occasion des échanges, dans des vocabulaires spécialisés, l'un des enjeux majeurs de cette maîtrise de l'information est de pouvoir repérer et manipuler les concepts désignés grâce aux vocabulaires spécialisés. Cette question a été l'une des préoccupations centrales de la documentation, notamment lorsque celle-ci s'est informatisée. Dans la mesure, en effet, où le mode de représentation le plus courant des contenus était le *descripteur* (voir la définition des termes *descripteur* et *mot-clé* en 3.3.1), la question s'est vite posée de trouver les moyens d'automatiser son repérage par des moyens mécaniques. Dans un deuxième temps, l'émergence de la linguistique comme discipline dominante pour l'étude des phénomènes d'expression suscite l'espoir de pouvoir – à travers l'étude des structures de surface (morphologiques et syntaxiques) – dégager les concepts porteurs de sens. Dès lors, commence une troisième phase qui voit la terminologie se pencher sur ce problème, confrontée, en effet, à la nécessité de trouver des méthodes de repérage de « mot-clé » ou *descripteur* qui prennent en charge l'environnement du mot-clé.

2. TERMINOLOGIE ET ACCÈS À L'INFORMATION

La terminologie est perçue comme une représentation du contenu de documents et comme une clé de l'accès à leur contenu. La création des ressources terminolo-

logiques nécessaires à ce double processus implique le recours à une vaste documentation. Aussi, dès l'apparition de la terminologie en tant que discipline autonome, le lien qu'elle avait avec la documentation a-t-il été vu comme essentiel. La diversité des contextes dans lesquels le travail terminologique est réalisé, la diversité des points de vue selon lesquels elle est employée et la variété des sujets de la terminologie, qu'il s'agisse des thèmes qu'elle traite ou des finalités qu'elle poursuit donnent lieu à un large éventail d'approches et d'applications qui se situent dans le champ des sciences de l'information et de la communication (Mustafa El Hadi 2004).

Selon les positions théoriques de la conception historique, la spécificité de la terminologie ne peut se définir que dans le modèle de communication. L'activité terminologique dès les débuts visait non seulement à normaliser le vocabulaire employé dans un domaine bien défini, mais aussi à faciliter la communication en élaborant des outils de référence pour des domaines spécialisés (Cabré 1998). La terminologie intervient ainsi dans l'élaboration de vocabulaires de référence pour des applications de plus en plus diversifiées en accès à l'information : thésaurus pour les systèmes d'indexation automatique, index structurés pour la documentation technique hypertextuelle, référentiels terminologiques pour les systèmes de gestion de données techniques, ontologies pour les mémoires d'entreprise, terminologies pour les systèmes d'aide à la décision ou pour les systèmes d'extraction d'informations.

Le traitement et la représentation des documents doivent se baser sur une analyse adéquate du contenu textuel. Cela permet aux usagers des systèmes d'information de se faire une idée exacte des connaissances disponibles dans une source donnée et donc de l'utilité potentielle de cette dernière pour leurs besoins. Le défi actuel posé aux spécialistes de l'information, consiste précisément à trouver les meilleurs moyens d'extraire et de représenter ce contenu textuel.

3. L'INDEXATION

3.1. Indexation : définitions et historique

Pomart *et alii* (1997 : 284-287) définissent l'indexation comme étant une opération ayant « pour but de faciliter l'accès au contenu de documents ou d'un ensemble de documents à partir d'un sujet ou d'une combinaison de sujets (ou toutes autres entrées utiles à la recherche). Cela s'applique aussi bien à l'élaboration d'index situés généralement en fin d'ouvrage qu'à l'usage des langages documentaires pour analyser le contenu d'une collection de documents et permettre par la suite, grâce aux fichiers où à la banque de données ainsi alimentée, la recherche des informations répondant à une préoccupation particulière. Ainsi la notion d'indexation se trouve-t-elle au cœur de la pratique documentaire : elle constitue l'opération centrale du traitement des documents au sein de la chaîne documentaire classique (répérage, et acquisition des documents ; traitement et exploitation ; recherche et diffusion) ».

L'utilisation des index remonterait d'après Nie (2003) au XV^e siècle peu après l'invention de l'imprimerie. Les index (ou termes d'indexation) jouent un rôle important dans la recherche d'information dans la mesure où ils déterminent avec quels mots on peut retrouver un document. Le problème du choix des éléments de l'index s'est vite posé. Chez les praticiens et chercheurs de cette spécialité, la dialectique du choix entre une indexation manuelle ou automatisée, un vocabulaire libre ou vocabulaire contrôlé, ou encore la détermination d'un anti-dictionnaire (*stoplist*) polarisent encore les débats et les interrogations.

L'indexation est une forme de reformulation : indexer, c'est reformuler le contenu d'un document ou d'une inscription dans une forme plus adaptée à son contexte d'exploitation dans une application donnée. On indexe donc en vue d'une application. On ne parle plus seulement d'indexation mais également d'enrichissement, d'annotation, de marquage et de métadonnées (Menon 2004).

3.2. Les techniques d'indexation

L'indexation se situe dans un contexte plus global, qui est celui de l'analyse du contenu. Cette opération est un préalable indispensable à toute recherche d'information sur le contenu et à d'autres types de traitement des informations. Il est donc difficile de parler de *recherche d'information* sans parler d'*indexation*, au sens procédural du terme. *Indexation* et *recherche d'information* sont de ce fait deux processus fortement liés. On indexe les documents dans le but de les retrouver. Le processus d'indexation est plus ou moins complexe suivant la représentation des unités documentaires ou de contenu que l'on cherche à atteindre, la disponibilité de ressources, comme les référentiels terminologiques (liste contrôlée de termes, thesaurus, classifications, ontologies de domaines) et l'application visée.

3.2.1. L'indexation manuelle

L'indexation manuelle est une opération qui consiste à recenser les concepts dont traite un document et à les représenter à l'aide d'un langage documentaire (classificateur ou combinatoire) souvent par plusieurs indicateurs sémantiques : indice de classification, descripteurs libres, termes autorisés, descripteurs ou mots-clés d'un thesaurus (Maniez 2002, Lefèvre 2000).

L'indexation manuelle fait recours à des outils d'indexation censés réduire la subjectivité des indexeurs, c'est-à-dire des « langages documentaires » (voir la définition en 3.3.1.). Ces outils servent à maintenir la cohérence et l'homogénéité. Ce type d'indexation repose habituellement sur un jugement de signification plus ou moins intuitif, toujours lié à l'indexeur. Le travail à réaliser pour la mise au point d'une indexation est assez important : connaissance du contenu de l'information, choix des concepts à représenter et traduction de ces concepts en descripteurs. De plus, les mêmes notions peuvent être exprimées de manières très diverses.

On a donc assez vite pensé à avoir recours à l'ordinateur :

- pour limiter les choix parfois subjectifs de l'indexeur ;
- pour alléger le travail requis par une indexation manuelle ;

- pour éviter les incohérences résultant des interprétations différentes entre plusieurs indexeurs ;
- et pour réaliser une recherche exhaustive des sujets traités dans l'information analysée.

Malgré ses limites, cette indexation « humaine » reste le seul type d'indexation intellectuelle (faisant appel à des connaissances sémantiques). Néanmoins, l'utilisateur est absent de cette indexation qui se fait en amont (les listes de concepts et des mots-clés sont constitués en amont donc, indépendamment des documents à indexer). Par ailleurs, les listes de concepts ou de mots-clés peuvent être mis à jour en fonction de l'évolution des connaissances mais dans la pratique cette mise à jour représente une lourde tâche qui n'est pas souvent effectuée en temps réel. Ceci laisse peu de place à l'évolution du vocabulaire de l'indexation en fonction de celle des documents indexés pris isolément. En raison de ces divers problèmes, des méthodes d'indexation automatique sont donc apparues.

3.2.2. *L'indexation automatique*

L'indexation automatique est l'opération qui consiste à faire reconnaître par l'ordinateur des termes figurant dans le titre, le résumé, le texte complet (s'il est enregistré avec la notice documentaire) et à employer ces termes, soit tels quels soit après conversion en d'autres termes équivalents ou conceptuellement voisins, pour en faire des critères incorporés dans le fichier de recherche et utilisables pour retrouver le document. La première génération des systèmes d'indexation automatique sur le texte intégral était fondée sur la création d'index (fichiers inversés) permettant de localiser un terme ou un ensemble de termes au sein d'une collection de documents.

La première approche d'indexation automatique KWIC (Key-Word In Context) a été introduite par Luhn (1957). Il a été question ensuite de pondérer les index. Aux débuts de la recherche d'information, les méthodes statistiques ont été fondées sur la fréquence de termes dans le document. Plus tard, cette mesure a été étendue pour tenir compte de la spécificité d'un terme pour le document. À cette fin, d'autres méthodes ont été exploitées, telles que *2-Poisson* (Nie, 2003).

L'indexation automatique a comme base :

- des méthodes statistiques et probabilistes, qui, pour sélectionner des termes d'index, combinent critères distributionnels (fréquence, cooccurrence de mots), positionnels (la partie du document dans laquelle un mot apparaît : titre, corps du texte, sommaire) et surtout quantitatifs (longueur du texte considéré) ;
- des méthodes linguistiques : lexicographiques et morphosyntaxiques. Ce sont les techniques employées dans le traitement automatique des langues naturelles ;
- des méthodes informatiques (telles que : algorithmes de recherche, langages évolués spécifiques, etc.) utilisées aussi bien dans le traitement automatique des langues qu'en documentation automatique.

Les systèmes statistiques considèrent les documents comme de simples chaînes de caractères et de ce fait indexent les mots isolés. La nature linguistique des textes n'est pas prise en compte, néanmoins, diverses expériences ont montré que l'amélioration de la performance des systèmes d'indexation passe par l'intégration d'au moins deux modèles – linguistique et statistique.

3.3. Place de la terminologie

Le terme, objet d'étude de la terminologie, est le fondement même des activités documentaires et informationnelles classiques comme l'indexation, la construction de thesaurus, la recherche d'information, ou plus récemment la veille, la gestion de contenu et la gestion de connaissances. La pratique terminologique peut être abordée à travers ses réalisations concrètes (index, thesaurus, ontologies, réseaux sémantiques...) comme un dispositif de médiation des savoirs et des connaissances. Le terme constitue ainsi un objet d'étude à part entière pour la terminologie mais aussi pour la documentation à travers l'étude des langages documentaires (voir la définition en 3.3.1). Le lien entre terminologie et documentation se vérifie ainsi à travers les langages d'indexation fondés sur les thesaurus dont la construction fait appel aux ressources terminologiques.

L'organisation des connaissances a longtemps été réalisée par les grands systèmes classificatoires et les encyclopédies. L'accès à l'information a nécessité le recours à d'autres types d'outils tels que les listes d'autorité, les thesaurus, les index structurés et les taxonomies. De nos jours et dans le cadre de la recherche d'information sur le Web, d'autres dispositifs terminologiques viennent compléter ce panorama d'outils. Il s'agit de réseaux sémantiques, de cartographie conceptuelle et d'ontologies qui sont issus des développements en intelligence artificielle et plus récemment en ingénierie des connaissances. Ce sont des structures d'organisation et de représentation des connaissances innovantes qui font appel aux technologies du Web. Conçues sur la base de la modélisation conceptuelle d'un domaine de connaissance, ces structures sont de plus en plus utilisées dans des applications de gestion des connaissances, de veille technologique, d'indexation et de recherche d'information. C'est surtout dans les fonctions d'indexation et de recherche d'information que les systèmes d'organisation des connaissances classiques et innovants convergent vers des points communs.

Thesaurus, cartes conceptuelles et ontologie, partagent des caractéristiques communes, du fait que tous les trois sont des référentiels terminologiques conçus sous forme de réseau sémantique et sont utilisés en représentation du contenu de documents et en accès à l'information.

3.3.1. Le descripteur, élément central en indexation

L'indexation d'un document consiste en la caractérisation de son contenu à l'aide de descripteurs afin de pouvoir retrouver aussi facilement que possible dans l'ensemble de la masse documentaire les informations pertinentes qu'il contient. Pour définir le *descripteur* on pourra dire qu'il y a identité entre cet élément et le *terme*. Ce dernier est l'élément commun à la terminologie et aux systèmes

d'information. Le descripteur est l'unité de représentation d'un contenu qui est de nature conceptuelle. Il est considéré comme clé d'accès aux documents dont il synthétise le contenu. Cette définition correspond à l'approche normative qui met l'accent sur une symétrie entre *indexation* et *recherche d'information*, comme le souligne Amar (2000). Les termes *descripteurs* et *mots-clés* appartiennent plutôt au domaine de l'indexation. Dans le cadre de l'indexation automatique, un descripteur est par exemple un syntagme nominal pondéré que l'on juge apte à caractériser le contenu d'un texte. En indexation manuelle, il est un élément de thesaurus. Il dispose dans ce dernier cas de relations conceptuelles propres au thesaurus (générique, spécifique, associé, synonymique). Un mot-clé, souvent employé dans le cadre de l'indexation manuelle, est un élément du lexique, il appartient donc à une liste non structurée de vocabulaire d'indexation. Les descripteurs provenant du thesaurus et les mots-clés issus d'un lexique sont employés dans le cadre de l'indexation manuelle. En revanche, un *terme* au sens terminologique est une unité textuelle qui renvoie à des notions dans un domaine de connaissance déterminé. De ce fait, le *terme* relève de la nomenclature établie d'un domaine. Un *terme*, comme nous l'avons déjà évoqué, peut être un descripteur s'il est retenu par un procédé automatique de pondération ou s'il provient d'un thesaurus ou d'un lexique.

Les descripteurs peuvent appartenir à deux types de langages : le langage libre ou le langage contrôlé. Un *langage documentaire* est une liste contrôlée de termes d'indexation ayant fait l'objet d'une validation humaine (Jacquemin 1997).

Le langage libre est un langage évolutif, proche de la langue naturelle. Son vocabulaire, l'ensemble des éléments qui composent le langage, est choisi *a posteriori* dans les textes à indexer, et n'est donc pas limité par un contrôle extérieur au système de recherche d'information. Dans ce cas précis, les éléments du langage sont principalement des mots du texte et non des termes.

Le langage contrôlé est équivalent au langage documentaire. Il est formé d'un vocabulaire préétabli et d'une série de règles traduisant différentes relations possibles entre ses éléments. Construit *a priori*, ce langage doit être connu avant d'indexer les documents. En général, les langages contrôlés sont composés des termes décrits dans un thesaurus ou un lexique (Roussy *et alii* 1999). L'étude de la sélection des unités est partiellement prise en charge par la terminologie.

4. L'ÉVOLUTION DES TECHNIQUES EN ACQUISITION DE RESSOURCES TERMINOLOGIQUES ET SON IMPACT SUR LA REPRÉSENTATION DU CONTENU DE DOCUMENTS

L'extraction terminologique suppose deux activités essentielles : l'acquisition terminologique (dans un domaine précis, par exemple) qui implique la découverte de nouveaux termes, et l'indexation. Le but de l'acquisition terminologique est la constitution de ressources terminologiques telles que les vocabulaires contrôlés (thesaurus), index structurés, ontologies. Les thesaurus sont utilisés pour assister le processus de recherche d'information lors de la phase d'expansion de la requête pour la recherche d'information interlinguale ou pour l'interrogation de bases de

données. La constitution automatique de thesaurus est une tradition déjà ancienne et les différentes études menées dans ce domaine concernaient soit la découverte de nouveaux termes ou l'établissement de relations sémantiques (*cf.* Jacquemin 1997 pour de plus amples détails). Les outils d'extraction terminologique bilingues (Dagan *et alii* 1994), fondés sur une analyse des parties du discours et des technologies d'alignement visant à extraire des termes candidats et leurs traductions peuvent être utilisés pour ces deux tâches. Les composants monolingues de ce type de systèmes peuvent être utilisés pour l'indexation et pour l'extraction terminologique alors que les composants bilingues sont utilisés pour la traduction ainsi que pour la recherche d'information dans des bases documentaires multilingues (*ibid.*). Bien que ce type de système ait été conçu initialement pour l'extraction et la traduction de termes, il a été ensuite utilisé également et adapté pour l'indexation, comme l'expliquent Landauer *et alii* (1990), cité par Dagan *et alii* (1994).

Les deux facteurs les plus décisifs qui ont eu un impact réel sur l'usage de la terminologie et son insertion dans les processus d'indexation et de recherche d'information sont la mise à disposition de grande quantité de textes et la prise en compte de la dimension de l'usage. Concernant ce dernier facteur, de nombreux auteurs soulignent l'importance du fonctionnement réel du terme au sein du discours et insistent sur la prise en charge de la dimension de l'usage. La question de l'usage est au cœur de la problématique d'indexation et de recherche d'information.

4.1. Les autres unités dans l'environnement du terme et leur fonction

Les textes spécialisés ne contiennent pas que des *termes*. On peut y trouver aussi des mots. Il est nécessaire d'examiner le statut de ces unités. Comme nous l'avons déjà signalé dans nos travaux antérieurs (1992) et récents (2004), les mots de la langue ne sont pas « aptes » à dénoter directement l'univers de référence. Autrement dit, le lexique donne le sens du « mot », indépendamment de la référence, la terminologie s'occupe du « mot » dans un référentiel donné. Il convient de signaler que les besoins actuels ont amené les terminologues à s'intéresser aux autres unités qu'on appelait « mots de la langue ». Les unités qui intéressent les terminologues pour des applications de traduction et de rédaction techniques sont essentiellement de deux catégories : les verbes et les adjectifs. L'Homme (2001) a signalé l'importance de prendre en considération la présence de verbes chargés d'informations terminologiques. De nombreux chercheurs s'intéressent à l'environnement du terme et construisent des outils ayant comme objectif l'extraction de collocations verbales et adjectivales, unités nécessaires en rédaction et en traduction humaine et/ou automatique. La collecte des données sur les termes a été facilitée par les concordanciers qui extraient rapidement des contextes associés à un terme donné à la demande d'un utilisateur (L'Homme, *ibid.*). De nouveaux outils qui permettent d'isoler des contextes spécifiques ou d'exploiter l'information qui se trouve de part et d'autre des termes préalablement recherchés ont fait leur apparition.

D'autres chercheurs vont plus loin et proposent ce que L'Homme (*ibid.*) appelle une « Nouvelle génération de dictionnaires spécialisés ». D'après elle, les dictionnaires classiques décrivent les termes de nature nominale dans des fichiers ou des articles monoconceptuels et retiennent les rubriques suivantes : des vedettes dans une ou plusieurs langues, une définition, des contextes et des marques d'usage. On assiste de plus en plus à des changements qui touchent la nature des données : les supports de données terminologiques tentent d'intégrer des éléments descriptifs différents et plus riches que ce que l'on trouve dans les dictionnaires terminologiques élaborés selon les règles de l'art. Ils tiennent compte de la polysémie des unités appartenant à des catégories différentes de celles du nom et peuvent retenir des rubriques comme les co-occurents ou les termes apparentés morphologiquement à l'unité terminologique (voir comme exemple de ce type de dictionnaire le *Dictionnaire d'apprentissage du français des affaires* (DFA), (Binon *et alii* 2000, cité par L'Homme, *ibid.*) comme étant le dictionnaire le plus riche sur le plan linguistique).

Pour sa part, Dancette (2005) signale l'apport similaire que constituent les outils d'interrogation (dictionnaires), qui ne portent plus principalement sur les termes, mais aussi sur leurs contextes (phraséologies, collocations) et sur les concepts. L'utilité de cette approche pour la recherche documentaire et terminologique paraît importante. Cette « révolution dictionnaire », précise l'auteur, résulte d'une convergence des disciplines : lexicologie, terminologie et traductologie. L'auteur et son équipe, à l'Université de Montréal, proposent un modèle de relations lexicosémantiques (RLS) permettant, à l'intérieur d'un domaine de spécialité (en l'occurrence le domaine commercial), de présenter le vocabulaire, en mettant en évidence ce qui relie un terme à un ensemble d'autres termes du même champ sémantique. L'idée s'inspire de plusieurs courants : la lexicologie explicative et combinatoire, les ontologies utilisées en terminologie et les recherches sur les collocations dans les domaines spécialisés. Les liens que cette démarche cherche à formaliser sont conceptuels (cause, instrument, but, agent, etc.), ou ontologiques (hiérarchie, partie-tout, synonymes, antonymes, etc.). Ils interviennent sur des plans différents : paradigmatique, syntagmatique et dérivationnel.

On voit apparaître de plus en de plus de technologies qui ont pour but l'amélioration de l'environnement terminologique et de la qualité de l'extraction. Des études relativement récentes dans ce domaine concernent essentiellement les collocations et les verbes de spécialité.

Pour améliorer la qualité et la pertinence de l'extraction, tout particulièrement dans le domaine de l'indexation, la rédaction technique et la traduction, l'identification et la recherche de ces éléments est devenu un objet d'intérêt primordial pour les spécialistes du TAL appliqué à la terminologie et à l'extraction terminologique (*cf.* L'Homme 2001, 2002 ; Fabre *et alii* 2002 ; Daille 2001). Concernant les adjectifs relationnels inclus dans les noms composés, une étude intéressante a été menée par Daille (*ibid.*). Cette étude montre comment les adjectifs peuvent être utilisés dans différents types d'applications du TAL. Dans le contexte des ressources terminologiques pour la recherche d'information, la méthode suggérée par cet auteur a pour but de regrouper les formes synonymes renvoyant à un concept unique. Cette mé-

thode est utilisée pour la mise à jour des thesauri. C'est d'ailleurs dans cet objectif de mise à jour que des logiciels développés pour l'extraction automatique de termes étaient destinés à enrichir les thesaurus. Lexter, par exemple, a été créé pour extraire les nouveaux termes ou termes émergents dans le domaine électrique à mettre dans l'index d'EDF-GDF (*cf.* Bourigault 1992).

De plus, l'extraction de collocations fait apparaître des affinités entre des termes et d'autres mots, des collocations ou d'autres co-occurents. Ces derniers peuvent servir d'indice pour dégager des « familles conceptuelles », (Dancette 2005 ; L'Homme 2001). Dans certains programmes, les cooccurrents servent d'indicateurs sur la parenté conceptuelle des termes et sont utilisés pour regrouper des ensembles de termes sous forme de classes. On se fonde sur le fait que si des termes ont les mêmes cooccurrents, ils partagent des caractéristiques sémantiques communes ou appartiennent à la même classe. Ce fait constitue un réel intérêt dans les contextes d'indexation et de recherche d'information.

5. CONCLUSION

Pour conclure, de nombreux travaux (L'Homme 2001, 2002 ; Dancette 2005 ; Bourigault *et alii* 2002 ; Daille 2001 entre autres), montrent que désormais le terminologue a accès à toutes sortes de techniques pour repérer et extraire des informations présentes sous différentes formes dans les textes spécialisés : extraction de termes complexes, termes simples, de relations sémantiques ; repérage de collocations et de contextes ciblés ; établissement de liens conceptuels entre termes ou construction de classes. Tous ces éléments représentent un apport essentiel dans nos pratiques informationnelles telles que la représentation du contenu de documents et la construction de référentiels terminologiques pour l'accès à l'information.

BIBLIOGRAPHIE

- Amar (M.), 2000 : *Les Fondements théoriques de l'indexation : une approche linguistique*, ADBS, Paris.
- Bourigault (D.), 1992 : « Lexter, vers un outil linguistique d'aide à l'acquisition des connaissances », dans *Actes des 3^{es} Journées d'acquisition des connaissances*, Dourdan, avril 1992.
- Bourigault (D.), Lame (G.), 2002 : « Analyse distributionnelle et structuration de terminologie : application à la construction d'une ontologie documentaire du droit », dans Nazarenko (A.) et Hamon (Th.), 2002 *Structuration de terminologie* ; TAL, vol. XLIII– n° 1, Hermès, Paris, pp. 128-150.
- Cabré Castellvi (M.-T.), 1998 : *La terminologie — Théorie, Méthode et Applications*, traduit du catalan, adapté et mis à jour par M.-C. Cormier et J. Humbley, Paris/Montréal, Armand Colin/Les Presses de l'Université de Montréal, 322 pp.
- Condamines (A.), Gilles-Aussenac (N.) édité, 2003 : « ASSTICCOT, Rapport Action Spécifique STIC « Corpus et Terminologie » (AS 34), rapport interne, n° IRIT/2003-23-R.
- Dagan (I.), Church (K.), 1994 : « Termight : Identifying and Translating Technical Terminology », dans *Proceedings of 4th Applied NLP Conference*, pp. 34-40.
- Daille (B.), 2001 : « Qualitative Terminology Extraction : Identifying Relational Adjectives », dans *Recent Advances in Computational Terminology*, Bourigault (D.), Jacquemin (C.), L'Homme (M.-C.), édité, Amsterdam, John Benjamins, pp. 149-166.

- Dancette (J.), 2005 : « Les représentations lexico-sémantiques ; moyen de structuration de connaissances dans les domaines spécialisés », dans *Actes du quatrième congrès d'ISKO-France*, « L'organisation des connaissances : approches conceptuelles », Grenoble, 3-4 juillet 2003, L'Harmattan, coll. « La Librairie des humanités », 2005, pp. 83-93.
- Fabre (C.), Frérot (C.), 2002 : « Groupes prépositionnels arguments ou circonstants : vers un repérage en corpus », dans *Actes du colloque TALN 2002*, Nancy, I, pp. 215-224.
- Jacquemin (C.), 1997 : *Variation terminologique : reconnaissance et acquisition automatiques de termes et de leurs variantes en corpus*, mémoire d'habilitation à diriger des recherches en informatique fondamentale, Université de Nantes.
- Lefèvre (P.), 2000 : *La Recherche d'information, du texte intégral au thésaurus*, Paris, Hermès, 253 pp.
- L'Homme (M.-C.), 2001 : « Nouvelles technologies et recherche terminologique : techniques d'extraction des données terminologiques et leur impact sur le travail du terminographe » dans *Actes du colloque « L'impact des nouvelles technologies sur la gestion terminologique »* collège universitaire Glendon de l'Université York, Toronto, 18 août 2001 (www.onterm.gov.on.ca/onterm/iso/proceedings.html).
- L'Homme (M.-C.), 2002 : « Fonctions lexicales pour représenter les relations sémantiques entre termes », Nazarenko (A.), Hamon (T.), (sous la dir.), TAL, vol. XLIII, n° 1, *Structuration de terminologie*, Paris, Hermès, pp. 19-41.
- Luhn (H.-P.), 1957 : « A Statistical Approach to Mechanized Encoding and Searching of Literary Information », dans *IBM Journal of Research and Development*, vol. LII, n° 12, pp. 309-317.
- Maniez (J.), 2002 : *Actualité des langages documentaires ; Fondements théoriques de la recherche d'information*, ABDS, Paris, 2002.
- Menon, (B.), 2004 : « L'indexation à l'heure du numérique », Journée d'étude ADBS, dans *Documentaliste – Sciences de l'information*, 2004, vol. XLI, n° 6.
- Mustafa El Hadi (W.), 1992 : « La contribution de la terminologie à la conception théorique des langages documentaires et à l'indexation de documents », dans *Metα*, 37 (3), Montréal, 1992, pp. 465-473.
- Mustafa El Hadi (W.), 2004 : *Terminologie et accès à l'information : Les nouveaux besoins et leur impact sur l'évolution des deux disciplines*, mémoire d'habilitation à diriger les recherches en science de l'information et de la communication, Université de Lille III, novembre 2004.
- Nie (J.-Y.), 2003 : « Introduction. Le domaine de la recherche d'information, survol d'une longue histoire » dans Gaussier (E.), Stefanini (M.-H.) (dir.) (2003), *Assistance intelligente à la recherche d'information*, collection « Traité des sciences et techniques de l'information », Paris, Lavoisier, pp. 19-28.
- Pomart (P. D.), Sutter (E.), 1997 : « Indexation », article du *Dictionnaire encyclopédique de l'information et de la documentation*, Cacaly (S.), le Coadic (Y.-F.), Melot (M.), Pomart (P.-D.), Sutter (E.), Paris, Nathan, pp. 284-287.
- Roussy (C.), Calabretto (S.), Pinon (J.-M.), 1999 : « État de l'art en indexation et recherche d'information », dans Dupoirier (G.), Ermine (J.-L.), (édit.) (1999) : *Gestion des documents et gestion des connaissances*, Document numérique, vol. III, n° 3-4.

MARQUEUR, UN LOGICIEL DE MARQUAGE SEMI-AUTOMATIQUE DE TEXTES

Abdelkrim Mokhtari

département de langue et de littérature françaises

Université Ibn Tofail

faculté des lettres et des sciences humaines

Kénitra, Maroc

mabelkrim@hotmail.com

Résumé : Marqueur est un logiciel de marquage semi-automatique des textes. Il sert à attribuer aux mots le lemme et la catégorie appropriés au contexte. Marqueur utilise un dictionnaire de contextes qui l'aide à résoudre certains cas ambigus. Pour les cas problématiques, Marqueur offre une interface de traitement interactif. Il offre aussi la possibilité de marquer les mots par d'autres marques (note, objet multimédia, etc.). Un accès libre au dictionnaire et au texte fait de ce logiciel un outil d'exploration et de manipulation des données textuelles.

Mots-clés : marquage, étiquetage, annotation de textes, TAL, exploration de textes, dictionnaire électronique

Nous nous proposons dans cet article de présenter un outil informatique qui s'inscrit dans le cadre général de la linguistique informatique ou, dit autrement, du traitement automatique des langues (TAL). Cet outil, baptisé *Marqueur*¹, est un logiciel de marquage semi-automatique de textes. Il a été conçu et développé par Abdelkrim Mokhtari (Université Ibn-Tofail, Kénitra, Maroc) et Michael Mepham (Université Laval, Québec, Canada). Prévu initialement pour permettre le marquage au niveau du lemme et de la catégorie, ses capacités ont été étendues à l'introduction d'autres marques, appelées *marques ouvertes*, comme à diverses tâches relatives à la manipulation des données textuelles. Bien que son développement soit déjà avancé, ce programme n'est pas tout à fait achevé et d'autres idées peuvent être prises en compte.

Marqueur est un programme ouvert et peut être utilisé pour accomplir différentes tâches relatives à l'analyse ou à l'exploration de textes. Le public ciblé est très large : lexicographes, terminologues, analystes du discours ou tout autre chercheur

¹ *Marqueur* a été développé dans le cadre d'une thèse (Ph. D.) en linguistique informatique (voir bibliographie). Abdelkrim Mokhtari a bénéficié durant l'été 1996 d'une bourse d'échanges entre départements accordée par L'AUPELF-(FICU) pour améliorer la version initiale de *Marqueur*. Le logiciel a aussi été utilisé dans un projet soutenu par L'AUPELF sur la désambiguïsation lexicale et les recherches textuelles. Il fait également l'objet de mémoires des étudiants de 4^e année à la faculté des lettres, Université Ibn Tofail, Kénitra, Maroc.

intéressé par la manipulation du lexique dans un texte donné. Plus précisément, *Marqueur* peut être utile à tout utilisateur qui a besoin de :

- marquer les mots d'un texte en leur assignant respectivement le lemme adéquat et la catégorie grammaticale adéquate ;
- attacher aux mots d'un texte des marques libres telles qu'une note, un signet ou un objet multimédia (image, son, clip vidéo etc.) ;
- faire des recherches sur le texte ;
- faire un commentaire de texte assisté par des outils d'exploration de texte ;
- constituer un dictionnaire sur la base de *corpus* variés.

Cette liste de tâches va nous servir de plan pour cette présentation.

1. MARQUAGE AU NIVEAU DU LEMME ET DE LA CATÉGORIE

La plupart des applications en TAL requièrent que le texte traité soit enrichi d'informations linguistiques. Qu'il s'agisse de traduction automatique, d'interprétation sémantique ou d'indexation, la connaissance par exemple du lemme ou de la catégorie mise en œuvre dans le contexte peut apporter un élément décisif. Cette information est d'autant plus importante que le nombre de mots qui posent un problème d'ambiguïté est élevé. *Marqueur* exploite un dictionnaire d'environ 250 000 formes dont environ 22 000 sont ambiguës quant au lemme ou la catégorie.

Marqueur, comme son nom l'indique, est un logiciel de marquage des textes. Il a été conçu pour assigner aux mots du texte d'abord le lemme et la catégorie. Si l'on tient compte des textes qui ont servi de test, cette assignation est accomplie de manière non problématique dans environ 80 % des cas. *Marqueur* utilise un dictionnaire répertoriant en principe toutes les formes du français. Chaque forme est enregistrée dans le dictionnaire soit comme univoque si on ne peut lui associer qu'un seul lemme et si elle ne peut être utilisée que sous une seule catégorie, soit comme ambiguë si plus d'un lemme peut lui être affecté ou si elle peut être utilisée sous plus d'une catégorie. Environ 81 % des formes du dictionnaire utilisé sont univoques. L'algorithme est donc simple. Passée la phase de segmentation, les mots relevés dans le texte (de manière séquentielle) sont confrontés aux formes du dictionnaire, s'il y a correspondance, l'information sur le statut de la forme est prise et le marquage s'effectue automatiquement.

Les formes univoques sont enregistrées dans le dictionnaire de manière unique. Deux homographes se verront inscrits comme une seule entité (« cher » au sens affectif et « cher » au sens de « onéreux » recevront une seule entrée) ; cette façon de faire ne rend pas justice à ceux qui prêchent la distinction des homonymes, mais elle apporte une simplification dans la représentation des informations lexicales. À l'inverse, deux homophones constitueront deux unités séparées si leur graphie diffère (« fard » et « phare »). Les mots à double graphie, même quand ils partagent la même signification, comme c'est le cas de « clé » et « clef », sont représentés en deux unités distinctes.

Les mots qui sont univoques sont marqués sur la base de l'information disponible dans le dictionnaire intégré dans le logiciel. Cette information étant unique, le marquage est par conséquent non problématique.

La représentation des lemmes présente un peu plus de difficultés. La logique suivie est de retenir toute entité discrète liée à la forme. Ainsi, la forme « avions » engendre-t-elle deux lemmes différents : « avoir » et « avion ». Mais faut-il retenir deux lemmes différents pour la forme « aïeux » (« aïeul » et « aïeux »), autrement dit, doit-on considérer que le lemme « aïeux » existe en tant qu'entité à part, ou bien a-t-on affaire seulement à une flexion du lemme « aïeul ». La question est encore plus sensible avec « nouvelle » qui peut engendrer deux lemmes « nouveau-nom » et « nouvelle-nom ». La forme « nouvelle » est à la fois une forme fléchie de « nouveau » et un lemme à part : « nouvelle-nom » au sens d'information.

Quoi qu'il en soit, nous avons organisé les données du dictionnaire de manière à permettre à l'utilisateur d'ajouter des lemmes ou d'en supprimer. Il peut également raisonner à un niveau supérieur au lemme et manipuler le concept d'hyper-lemme. Dans tous les cas, en revanche, le fichier des lemmes ne permet pas de retenir des doublons. Par ailleurs, l'utilisateur a, à tout moment, la possibilité de vérifier tous les lemmes pour une forme donnée et toutes les formes associées à un lemme donné.

Les mots qui sont ambigus sont traités selon un dictionnaire des contextes immédiats, dictionnaire également intégré dans le logiciel. S'il n'y a aucun contexte de disponible pour le mot envisagé, *Marqueur* ne fait rien et renvoie le traitement à l'étape du marquage interactif.

L'utilisation du dictionnaire des contextes repose sur l'hypothèse que certains contextes (seuls les contextes immédiats ont été implémentés dans *Marqueur*) sont suffisants pour lever l'ambiguïté lemmatique et catégorielle. Un exemple peut montrer l'intérêt de cette démarche.

[1] La valeur de son père en son temps sans pareille,
Tant qu'a duré sa *force*, a passé pour merveille (*Le Cid*)

Le mot « force » mis en italique dans le texte peut naturellement avoir deux lemmes différents : « forcer-verbe » et « force-nom ». La présence du mot « sa » immédiatement à gauche suffit à opter pour le second lemme.

On sera tenté d'étendre cette règle lexicale à tout le paradigme des adjectifs possessifs et même au paradigme des déterminants et d'envisager de retenir une règle du genre « déterminant » + force implique force-nom. Le même raisonnement tient pour le second exemple :

[2] Mais à ce peu de mots je crois que sa *pensée*
Entre vos deux amants n'est pas fort balancée. (*Le Cid*)

Le mot « pensée » qui doit être résolu en tant que « pensée-nom » et non en tant que « penser-verbe », peut bénéficier d'une règle de contexte grammatical du genre : déterminant + « pensée » implique « pensée-nom ».

Marqueur permet de retenir ces deux types de contextes et en tire profit à chaque session de marquage automatique.

L'utilisateur doit se convertir en linguiste expert et faire preuve de beaucoup de perspicacité pour ne pas retenir des règles contre-productives. Le terrain est miné et l'on doit soupeser toutes les règles que l'on veut établir. Cette démarche est inspirée des travaux d'Éric Brill (1995), mais elle ne recourt pas comme c'est le cas de la démarche de Brill à un *corpus* témoin, ni à un processus d'apprentissage de règles. Il s'agit pour nous d'alimenter le dictionnaire de contextes discriminatoires. Chacun de ces contextes aura fait l'objet d'une évaluation empirique. Le logiciel permet par ailleurs d'éliminer les cas les moins heureux.

Nous avons évoqué plus haut le marquage interactif. En effet, pour tous les cas non résolus, *Marqueur* offre la possibilité de les résoudre manuellement. L'interface a été conçue pour cependant disposer des informations appropriées contenues dans le dictionnaire. C'est une aide appréciable qui facilite l'analyse au sens où les différents lemmes ne sont pas forcément présents à l'esprit. C'est le *marquage assisté*. Si, par contre, l'utilisateur souhaite marquer un mot sans que ce mot ne soit enregistré dans le dictionnaire (on peut vouloir travailler avec un dictionnaire de spécialité et n'intégrer dans celui-ci que le lexique qu'on juge pertinent), il suffira de basculer vers le *marquage personnel*.

Le marquage assisté est cependant plus productif. Les mots nouveaux qu'on ajoute, les mots dont on rectifie les marques, les contextes qu'on retient, sont autant d'apports pour les analyses ultérieures.

Enfin, les mots absents du dictionnaire, à l'instar des mots ambigus non résolus, sont renvoyés à l'étape interactive, à l'exception de quelques cas qui sont traités d'après leur morphologie.

Mais *Marqueur* offre paradoxalement un avantage qui est généralement considéré comme un défaut dans le monde de l'automatisation : le logiciel permet de désactiver le recours aux contextes et aux règles (morphologiques pour l'instant) et ce pour permettre à l'utilisateur d'aboutir à un marquage exclusivement manuel. Ce qui pourrait être utile dans certaines situations. En fait, une interface appropriée permet de sélectionner les sources de marquage qu'on souhaite mettre en œuvre.

2. DES MARQUES OUVERTES

Marqueur sert aussi à affecter aux mots d'autres marques. Une note, une définition, un signet, mais aussi des informations de type multimédia (image, son, clip vidéo).

Ces marques peuvent être insérées soit dans le texte, soit dans le dictionnaire. Pour le premier cas, l'utilisateur peut juger pertinent d'attacher les marques appropriées au mot dans son contexte, sans laisser de traces dans le dictionnaire. On peut en effet être intéressé par l'apport de la mise en discours des mots. Des annotations de différents niveaux peuvent être retenues, recherchées plus tard. Pour le second cas, l'utilisateur jugera si l'information a une portée générale, si elle a la prétention à une valeur universelle. Elle sera enregistrée dans le dictionnaire, soit dans la partie

forme (la transcription phonétique par exemple peut être placée à cet endroit), soit dans la partie lemme (la signification peut trouver sa place à ce niveau). *Marqueur* rend possible au prix d'un paramétrage par l'utilisateur de renvoyer vers le texte les marques souhaitées à l'exclusion des autres, et inversement de renvoyer les marques spécifiques au texte vers le dictionnaire. Cette interaction entre les deux univers, l'univers du texte, donc du discours et de la création, et l'univers du dictionnaire, donc de la mémoire, est voulue dans la mesure où elle ajoute une grande souplesse dans la circulation et l'organisation des données linguistiques.

Une fois ces marques enregistrées, elles pourront servir dans les prochaines sessions de marquage. Les mots univoques se verront étiquetés automatiquement, sans autre forme de procès. Les mots ambigus devront attendre que l'ambiguïté soit levée.

3. FILTRAGE ET RECHERCHE AVANCÉE

Dans un souci de rendre la tâche de marquage interactif moins fastidieuse, mais aussi pour permettre à l'utilisateur de naviguer dans le texte selon des critères variés et pointus, nous avons enrichi l'interface de l'interactif par des outils de filtrage et de recherche avancée.

L'utilisateur peut ainsi sélectionner la liste des mots qui l'intéressent et les retenir comme candidats au marquage ou tout simplement les rendre visibles pour un examen à un niveau ou à un autre.

Le filtre est très puissant et permet de jouer sur la morphologie des mots que l'on veut retenir, sur leurs propriétés ou marques, sur le mode de traitement déjà subi. Une recherche avec des expressions rationnelles va être ajoutée.

Ainsi, on peut revoir des mots précédemment traités. Par exemple, on peut revenir sur les cas marqués comme adjectifs ou comme noms. Comme on peut sérier les cas qui ont bénéficié des règles contextuelles (lexicales ou grammaticales).

On peut même retenir uniquement les mots du texte qui font partie d'une liste de mots (préparée grâce aux services de *Marqueur*), une liste formée par exemple d'une thématique ou d'un champ lexical donné ou même des mots qui composent une œuvre entière. Cette souplesse de filtrage rend possible de tester des postulats linguistiques.

L'affichage des mots candidats au marquage se fait dans l'éditeur principal du traitement interactif. On passe d'un mot à l'autre en utilisant les outils de navigation. Mais on peut travailler sur n'importe quel mot visible dans l'éditeur ; il suffit de le rendre courant en double-cliquant dessus.

4. CONSTITUTION ET MANIPULATION DU DICTIONNAIRE

Marqueur permet de travailler sur un projet avec un dictionnaire vierge. Seule la structure du dictionnaire est imposée. La table des catégories est aussi laissée à la

discrétion de l'utilisateur. Bien entendu, les deux bases étant liées, on prendra soin de ne pas changer l'une sans se préoccuper de l'autre.

Le dictionnaire est modifiable. On peut le modifier à tout moment. On peut y accéder à travers le marquage interactif ou de manière directe à partir du menu général du logiciel. Le dictionnaire est structuré en formes et en lemmes respectifs, ce qui pourrait favoriser une étude du lexique. De leur côté, les contextes, rassemblés de manière structurée, peuvent être considérés comme un *corpus* spécialisé, tout à fait exploitable.

Des procédés de création ou d'alimentation automatique du dictionnaire sont en cours de développement. Il s'agit de permettre à l'utilisateur d'ajouter au dictionnaire les mots nouveaux épinglés dans un texte inédit.

Marqueur donne aussi les moyens de se constituer des glossaires particuliers. L'utilisateur doit pour ce faire envoyer le mot (et ses propriétés) dans un glossaire personnel.

5. COMMENTAIRE ASSISTÉ PAR DES OUTILS D'EXPLORATION

Il peut être commode de pouvoir commenter un texte, tout en tirant profit d'informations obtenues grâce à des outils d'exploration et de filtrage tant au niveau du texte que du dictionnaire. *Marqueur* offre cette possibilité et met à la disposition de l'utilisateur toute la base des contextes lexicaux et grammaticaux.

6. CONCLUSION

Compte tenu de ce qui a été dit, *Marqueur* se présente d'abord comme un outil de marquage semi-automatique de textes au service des terminologues, des lexicologues, et des développeurs d'application d'automatisation. S'il n'a pas la prétention d'apporter une solution forte au problème de l'étiquetage, il a au moins le mérite de permettre un marquage manuel souple et transparent.

Mais *Marqueur*, c'est également un outil ouvert à tous les chercheurs qui voudraient emmagasiner des connaissances pour les transférer ensuite dans des textes inédits. Les étudiants pourraient l'utiliser pour les mêmes raisons ou comme outil d'exploration de textes, voire d'extraction d'information. Enfin, *Marqueur* se présente comme un outil pédagogique, mettant à l'épreuve l'utilisateur pour une réflexion sur différents aspects linguistiques, tant au niveau du lexique qu'au niveau du fonctionnement du discours.

BIBLIOGRAPHIE

- Brill (E.), 1995 : « Transformation-Based Error-Driven Learning and Natural Language Processing : A case Study in Part-of-Speech Tagging », dans *Computational Linguistics*, volume XXI, numéro 4, pp. 543-565.
- Fuchs (C.), 1993 : *Les Traitements automatiques des langues*, Hachette, 304 pp.
- Fuchs (C.), 1996 : *Les Ambiguïtés du français*, Ophrys, 184 pp.

- Martin (W.), Heymans (R.) et Platteau (F.), 1988 : « Dilemma, an automatic lemmatizer » dans Willy Martin (éd.) *Computational Linguistics at the University of Antwerp*, Antwerp papers in Linguistics n° 56.
- Mepham (M.), 1989 : « Le Traitement interactif des documents » dans C. Ouellon, actes du coll. « La description des langues naturelles en vue d'applications informatiques », Québec, Université Laval.
- Mepham (M.), 1989 : « Lexitex : Textual Analysis Software » dans *Tools for Humanists*, Toronto, Center For Computing in the Humanities : U. Toronto, 49-51 pp.
- Mepham (M.), 1991 : « Efficient Pre-processing for the Creation of Large-scale Full-text Data Bases » dans *Actes du Colloque RIAO 91*.
- Mokhtari (A.), 1993 : « Cohésion et désambiguïsation lexicale automatique : cas de l'anaphore » dans *Actes des 7^e Journées de Linguistique*, Ciral, Université Laval, Québec.
- Mokhtari (A.), 1994 : « Collocateur : première version d'un logiciel de reconnaissance semi-automatique d'expressions » dans *Actes des 8^{es} Journées de Linguistique*, Ciral, Université Laval, Québec.
- Mokhtari (A.), 1998 : *Cohésion lexicale et automatisations : application à la désambiguïsation lexicale*. thèse Ph. D., Université Laval, Québec, Canada

**TOPOGRAPHIE BITEXTUELLE
ET APPROCHES QUANTITATIVES DE L'EXTRACTION
DE RESSOURCES TRADUCTIONNELLES
À PARTIR DE *CORPUS* PARALLÈLES**

Maria Zimina

SYLED-CLA²T

Université de la Sorbonne nouvelle, Paris III

Paris, France

zimina@msh-paris.fr

Résumé : *La notion de corpus parallèle comportant plusieurs volets qui correspondent chacun à une version d'un même texte dans deux ou plusieurs langues différentes, renvoie à des situations connues de coexistence de textes présentant des liens forts dans leur structuration. Les méthodes quantitatives permettent d'accéder à de nouvelles dimensions d'analyse de ces corpus pluritextuels. Les perspectives ouvertes par cette approche offrent aux traducteurs, terminologues, lexicographes, etc., des moyens automatisés pour explorer la structure des correspondances lexicales par la navigation textométrique en corpus.*

Mots-clés : *bitexte, corpus parallèles, statistique textuelle, textométrie, topographie textuelle*

1. APPROCHE TEXTOMÉTRIQUE DE *CORPUS* MULTILINGUES

Dans le contexte de l'informatisation de la société, chercheurs et praticiens sont confrontés à une croissance spectaculaire des *corpus* de textes multilingues provenant d'archives de textes traduits, de bases documentaires multilingues numérisées, de sites Web internationaux. Pour des raisons variées, diverses communautés s'intéressent aux données textuelles multilingues. Les historiens, les juristes, les terminologues analysent les *corpus* multilingues avec des outils d'exploration spécifiques permettant d'observer plus finement les correspondances entre différents volets de *corpus* comparés. Les informaticiens utilisent les ressources obtenues à partir de ces *corpus* pour améliorer les performances des outils de traduction automatique ou des moteurs de recherche pour le Web. Enfin, les ressources traductionnelles obtenues à partir des *corpus* multilingues sont utilisées avec profit pour les études menées dans le cadre de plusieurs disciplines des sciences du langage qui s'étendent de la linguistique contrastive à la lexicographie, de la traduction assistée par ordinateur à l'enseignement des langues étrangères, de l'analyse du discours à la linguistique computationnelle.

Pour extraire automatiquement des ressources à partir des données textuelles multilingues, il faut disposer d'outils de traitement automatique de *corpus* pluritextuels. Le développement de tels outils fait l'objet des recherches menées dans le domaine du traitement automatique des langues (TAL). Pour analyser des *corpus* de textes rédigés dans des langues différentes et en extraire des ressources réutilisables, les systèmes actuels de TAL font souvent appel à des savoirs *a priori* (linguistiques, pragmatiques, etc.). Les explorations appuyées sur les approches de statistique textuelle sont moins courantes.

Dans ce contexte, nous avons pu confirmer notre intuition que les connaissances des propriétés fréquentielles du matériau textuel acquises dans le domaine de la textométrie donnent accès à des savoir-faire précieux susceptibles d'enrichir les pratiques actuelles du traitement automatique de *corpus* multilingues (Zimina 2004b). La spécificité de la démarche textométrique réside dans le statut privilégié conféré aux données textuelles. Elle se caractérise par deux opérations principales qui visent à découper le texte en unités puis à comparer leurs distributions en les opposant les unes aux autres dans les différents contextes où ces unités apparaissent (Lebart et Salem 1994). Les interactions du domaine de la textométrie multilingue avec d'autres disciplines qui se situent à l'intersection de la linguistique et de l'informatique sont présentées sur la figure 1.

Textométrie multilingue

<http://www.cavi.univ-paris3.fr/lexicometrica/jadt/textometrie-multilingue/>

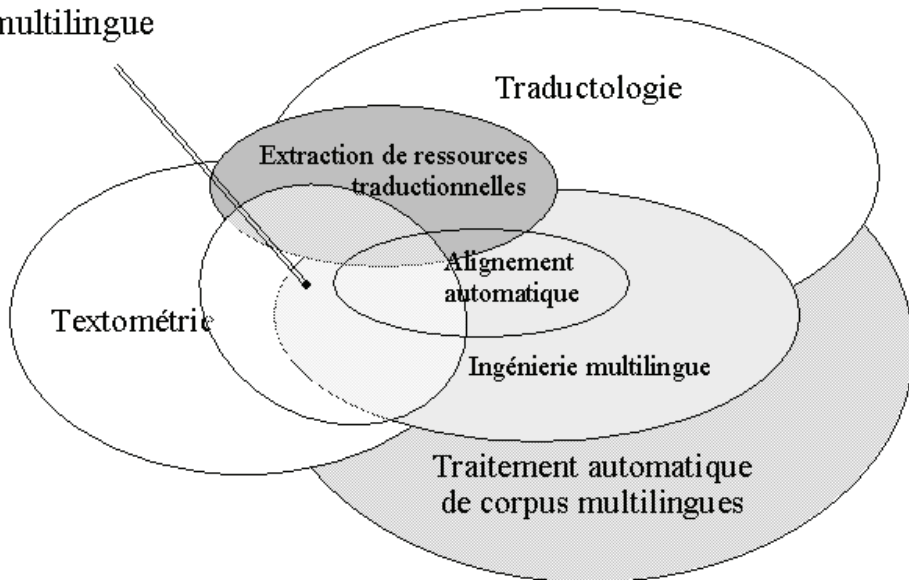


Figure 1 : Le domaine de la textométrie multilingue

2. UNITÉS DE TRADUCTION ET ANALYSE AUTOMATIQUE DE *CORPUS* PARALLÈLES

La textométrie multilingue fait appel à des méthodes quantitatives qui reposent entièrement sur des ressources construites automatiquement à base de *corpus*. Ces méthodes s'appuient sur des algorithmes qui utilisent les fréquences et les distributions des unités textuelles prises comme points de repère pour l'identification et l'extraction des correspondances. Les comparaisons des fréquences des unités textuelles recensées dans les différents volets d'un *corpus* pluritextuel sont souvent insuffisantes pour détecter les correspondances traductionnelles au niveau lexical. Les différents sens dans lesquels le lexème est employé dans un contexte donné induisent la plupart du temps autant de traductions différentes. Les mots dotés d'un large éventail de sens dans le *corpus* forment des réseaux de correspondances souvent complexes. Ces facteurs entraînent des écarts entre les fréquences des unités équivalentes prises dans des contextes particuliers.

2.1. *Corpus* parallèle français/anglais/russe « convention »

Pour explorer les phénomènes quantitatifs liés aux correspondances traductionnelles au niveau lexical, nous emprunterons des exemples à un *corpus* de textes juridiques français/anglais/russe de la *convention de sauvegarde des Droits de l'homme et des Libertés fondamentales*, désormais « convention ». Ce *corpus* a été constitué à partir du texte officiel de cette convention et de ses protocoles intégraux (français/anglais/russe)¹. Une série d'arrêts rendus par la Cour européenne des Droits de l'homme de Strasbourg est disponible en français (281 459 occurrences) et en anglais (274 475 occurrences). Le volet russe des arrêts de la Cour des Droits de l'homme est actuellement en cours de construction (cf. tableau 2).

Tableau 2 : La structure du *corpus* parallèle « convention »

<i>Corpus</i> convention	français	anglais	russe (traduction non officielle)
Convention de sauvegarde des Droits de l'homme	5 688 occurrences	5 375 occurrences	4 422 occurrences
Protocoles additionnels	5 411 occurrences	5 185 occurrences	4 066 occurrences
Arrêts de la Cour des Droits de l'homme	281 459 occurrences	274 475 occurrences	en cours de construction

Le tableau 3 présente les résultats de la segmentation automatique du *corpus* trilingue « convention » en occurrences des formes graphiques réalisée à l'aide d'outils de statistique textuelle *Lexico3* (Lamalle *et al.* 2003).

¹ Les textes français et anglais de la *convention de sauvegarde des Droits de l'homme* existent parallèlement ; il est difficile de distinguer une langue source et une langue cible. Le texte en russe est une traduction non officielle à partir de l'anglais. Cette traduction est disponible sur le portail du Conseil de l'Europe : <http://conventions.coe.int>

Tableau 3 : Résultats de la segmentation du *corpus* « Convention »

	occurrences	formes	hapax	Fréquence maximale	
français	11 099	1 438	666	<i>de</i>	690
anglais	10 560	1 226	529	<i>the</i>	1 003
russe	8 488	1 952	1 008	<i>в</i>	363
Liste de caractères-délimiteurs :., ; ! ?/_\`« » ()[]{}\$\$					

Nous constatons sur le tableau 3 que le nombre total d'occurrences est plus important pour le volet français (11 099 contre 10 560 pour l'anglais et 8 488 pour le russe). Le volet russe du *corpus* est le plus diversifié en terme de formes (1 952 contre 1 438 pour le français et 1 226 pour l'anglais) et compte beaucoup plus d'hapax que le français et l'anglais. Le plus grand nombre de formes dans le volet russe s'explique par la présence de désinences qui changent la forme graphique du mot (*cf.* tableau 4).

Tableau 4 : *Corpus* « Convention » : extraits des dictionnaires des formes

français	Fréquence	anglais	Fréquence	russe	Fréquence
article	170	article	169	статья	109
				статьи	41
				статье	5
				статьёй	16
} 171					
convention	110	convention	105	конвенция	11
				конвенции	77
				конвенцию	12
				конвенцией	9
} 109					
droits	51	rights	52	права	8
				прав	40
				правам	2
				правами	4
				правах	1
} 55					

L'analyse des dictionnaires des formes graphiques constitués à partir des trois volets du *corpus* permet de poursuivre la comparaison des caractéristiques fréquen-

tielles. On constate que pour une partie des formes en correspondance de traduction, la confrontation des dictionnaires classés par ordre lexicométrique² fait apparaître des similitudes entre leurs fréquences totales (F) dans le *corpus* (cf. tableau 4). L'analyse sémantique de ces équivalences lexicales en *corpus* permet de comprendre les causes de ces correspondances souvent univoques. Par exemple, le mot *article* est polysémique en français comme en anglais. Le mot équivalent en russe *статья* est également polysémique. Cependant, dans le *corpus* « convention », les mots *article* (F = 170)/*article* (F = 169)/*статья* (F = 171) (français/anglais/russe) sont employés dans les trois langues pour désigner la structure de découpage des documents juridiques. Le contexte du *corpus* réduit l'espace des sens possibles pour chacun d'entre eux :

français	anglais	russe
<u>article</u> 1 — obligation de respecter les droits de l'homme	<u>article</u> 1 — obligation to respect human rights	<u>статья</u> 1 — обязательство соблюдать права человека
la mort n'est pas considérée comme infligée en violation de cet <u>article</u> /.../	deprivation of life shall not be regarded as inflicted in contravention of this <u>article</u> /.../	лишение жизни не рассматривается как нарушение настоящей <u>статьи</u> /.../

Lorsqu'il s'agit de mots dotés d'un large éventail de sens dans le *corpus*, les correspondances lexicales entre les volets forment un réseau complexe et la comparaison des fréquences totales des formes graphiques ne constitue pas toujours une bonne indication pour l'appariement. Par exemple, la correspondance des mots polysémiques *droit* (F = 60)/*right* (F = 40)/*право* (F = 55) (français/anglais/russe) est seulement partielle. L'ancrage lexical se révèle indispensable pour cerner les variations contextuelles des traductions :

² Ordre lexicométrique : ordre résultant d'un tri des formes du *corpus* par ordre de fréquences décroissant ; les formes de la même fréquence sont classées par ordre lexicographique.

Français	Anglais	Russe
toute personne a droit à la liberté d'expression	everyone has the right to freedom of expression	каждый имеет право свободно выражать свое мнение
toute personne arrêtée ou détenue/.../ a le droit d'être jugée/.../	everyone arrested or detained/.../ shall be entitled to trial/.../	каждый задержанный или заключенный под стражу/.../ имеет право на судебное разбирательство/.../
le juge élu au titre d'un état partie au litige est membre de droit de la chambre	there shall sit as an ex officio member of the chamber/.../ the judge elected in respect of the state party concerned	судья, избранный от государства, являющегося стороной в деле, является ex officio членом палаты

2.2. Variations contextuelles d'équivalences traductionnelles

Nous avons montré que la comparaison des fréquences totales des formes isolées se révèle souvent insuffisante pour détecter des équivalences lorsqu'il s'agit des emplois polysémiques. L'étude des formes en contexte permet de lever certaines des ambiguïtés résiduelles ; au contact d'autres mots associés sur l'axe syntagmatique, différents composants du sens du mot sont activés et il devient possible d'en tenir compte lors de l'appariement. Nous allons ainsi examiner dans quelle mesure les comptages portant sur des unités textométriques plus larges, composées de plusieurs formes, sont susceptibles de fournir des indices pour l'extraction d'équivalences du *corpus*. Les résultats obtenus à partir de décomptes de formes graphiques isolées seront complétés par des comptages portant sur des objets textuels de type segment répété. Ces comptages peuvent être réalisés automatiquement à l'aide du logiciel *Lexico3* (Lamalle *et al.* 2003).

Le segment répété est une suite de formes dont la fréquence est supérieure ou égale à deux dans le *corpus*. Le terme « voisinage d'une occurrence » désigne tout segment qui contient cette occurrence dans le *corpus* (Salem 1987). Nous avons ainsi décrit le voisinage de la forme *état* dans le volet français du *corpus* par les segments répétés les plus fréquents ($F \geq 3$) recensés autour de cette forme (*cf.* tableau 5).

Tableau 5 : Corpus « convention » : voisinages lexicaux des pôles état/state/государство

F	Segment répété		
18	tout	état	volet français
11	un	état	
11		état peut	
9	tout	état peut	
8		état membre	
7	l'	état	
6	de l'	état peut au moment de la	
5	d'un	état	
5	tout	état peut au moment de la signature	
4	un	état membre du conseil de l'Europe ne	
4	tout	état peut à tout autre moment par la suite par	
4	tout	état peut au moment de la signature ou au moment	
4	pour	état membre qui exprimera ultérieurement son consentement à être	
4	territoire d'un	état	
3	par ledit	état	
3	un	état membre du conseil de l'Europe ne peut ratifier	
3	tout	état peut au moment de la signature	
3		état partie	
3	cet	état	
14	any	state	volet anglais
12	member	state	
11		state may	
10		state of	
9	the	state	
9	any	state may	
9		state may at	
8	any	state may at	
8		state which	
8	member	state of the council of europe	
7	of the	state	
6	a	state	
5	of a	state	
5		state may at the time of	
5	that	state	
4	the territory of a	state	
4	in respect of any member	state which subsequently expresses its consent	
4	a member	state of the council of europe may not ratify	
4	any	state may at any later date by a declaration addressed	
4	any	state may at the time of signature or when depositing	
4	by that	state	
4		state shall	
4		state which has made a declaration in accordance with paragraph	
3		state to	
11	любое	государство	volet russe
7		государство может	
5	любого	государства	
4		государства в	
4	любое	государство может	
4		государство при	
4		государство член совета европы	
3	на территорию	государства	
3	любое	государство при	
3	для любого	государства члена которое	
3	любое	государство при	
3		государство член совета европы не может ратифицировать принять или	
3	этого	государства	
3	генеральный секретарь совета европы направляет заверенную копию каждому	государству члену совета	

Dans le texte, ces récurrences d'unités ont des origines diverses (*cf.* tableau 5). Certaines d'entre elles correspondent à des syntagmes bien formés (p. ex. : *état membre, territoire d'un état*), d'autres résultent de la reprise partielle de fragments de phrases dont le statut syntaxique est intermédiaire (p. ex. : *pour tout état membre qui exprimera ultérieurement son consentement à être*). On remarque que les unités de ce type sont généralement dotées d'un sens que l'on ne peut pas déduire à partir des sens des mots isolés qui rentrent dans leur composition (p. ex. : *état partie*).

En dépit du statut intermédiaire de l'unité type segment répété du point de vue de la segmentation du *corpus* en unités de signification pertinentes, nous pensons que les inventaires de segments répétés recensés dans les volets parallèles du *corpus* constituent une ressource précieuse pour l'analyse des éléments formels d'expression qui servent à maintenir l'équivalence au niveau des mots et des syntagmes. En sélectionnant les segments trilingues français/anglais/russe dans lesquels sont attestées des formes équivalentes (p. ex. : *état/state/zocydapcmbo*), il devient possible d'obtenir un « effet visuel » qui met en valeur des ressources traductionnelles du *corpus* lors de l'édition automatique des listes (*cf.* tableau 5). Les constats produits par la méthode de segments répétés seront affinés par la « navigation textométrique » en *corpus* à l'aide de la topographie textuelle.

3. EXPLORATION TEXTOMÉTRIQUE DE TRADUCTIONS À L'AIDE DE LA TOPOGRAPHIE TEXTUELLE

3.1. Navigation textométrique intertextuelle

L'extraction des ressources traductionnelles à partir de *corpus* multilingues sur des bases quantitatives ne peut être menée en dehors d'une réflexion sur les problèmes de la segmentation parallèle de textes source et cible. Au cours de nos recherches, nous avons été confrontée à l'impossibilité de proposer un modèle de segmentation parallèle de *corpus* multilingues qui constitue directement des listes de correspondances traductionnelles et fournisse des représentations sémantiques et syntaxiques indépendantes du *corpus*. La détermination des mécanismes formels permettant d'automatiser cette segmentation nous semble d'ailleurs une perspective peu envisageable pour des raisons qui tiennent aux différences de structures des langues à tous les niveaux de l'analyse linguistique.

Ce constat a orienté notre travail vers la recherche de nouvelles stratégies d'analyse exploratoire de *corpus* multilingues. À côté des systèmes d'extraction automatique de lexèmes bilingues – candidats à l'appariement – présentés sous forme de listes figées, il nous paraît indispensable de fournir à l'utilisateur des outils de navigation dans les *corpus* multilingues, modulables en fonction de besoins particuliers. Dans cette perspective, l'utilisation de la « cartographie d'équivalences traductionnelles », fondée sur la représentation topographique du bitexte, nous semble particulièrement adaptée à l'analyse textométrique de *corpus* de traduction.

3.2. Cartographie d'équivalences traductionnelles

La topographie textuelle a pour objectif une localisation graphique des phénomènes mis en évidence par l'étude statistique (Lamalle et Salem 2002). Dans le contexte de cette approche, l'expertise humaine de textes est appuyée de multiples outils de lecture et visualisation qui offrent de nouveaux moyens d'investigation de l'espace textuel. Ainsi, l'étude de « l'organisation spatiale » d'occurrences d'équivalences lexicales peut être effectuée à partir de la description cartographique du bitexte.

Sur la figure 6, l'appariement des phrases des volets français et anglais du *corpus* « convention » est matérialisé par des carrés positionnés sur deux colonnes, la première correspond au texte en français et la deuxième au texte en anglais³. La construction de la représentation cartographique de l'alignement intègre le marquage de la partition « partie » qui distingue les documents juridiques faisant partie du *corpus*. Cette description cartographique du bitexte a été générée à l'aide de l'interface graphique du programme *mkAlign*⁴ qui permet de construire et visualiser l'alignement d'un couple de textes en modifiant au besoin la correspondance entre leurs segments respectifs.

Les fonctionnalités développées au sein de *mkAlign* permettent à l'utilisateur de visualiser la ventilation des unités textuelles à l'aide de la « carte des sections parallèles ». L'utilisateur dispose d'un ensemble d'outils permettant de choisir (à partir du dictionnaire des formes graphiques, de la liste importée des segments répétés, etc.) un type d'unité textuelle sur lequel portera son exploration. Après avoir sélectionné le type, il est possible d'étudier sa ventilation dans le *corpus* aligné. L'activation de l'unité choisie (par un clic droit) déclenche sa recherche et l'affichage des résultats de cette recherche sur la carte. La ventilation de l'unité est alors représentée sur la carte par des carrés « cochés ». Les carrés sont liés et toute sélection dans un volet est automatiquement répercutée dans l'autre. Le segment bitextuel lié à la section activée sur la carte s'affiche dans la fenêtre du bas de l'éditeur (cf. figures 6-7).

La carte des sections représentée sur la figure 6 montre la ventilation du segment français *tout état* (F = 18) et du segment anglais *any state* (F = 14). L'écart entre les fréquences totales de ces segments équivalents nous amène à nous interroger sur les variations contextuelles des traductions. La confrontation des deux volets de la carte révèle une correspondance presque totale dans la répartition des segments dans les phrases alignées du *corpus*. En faisant un clic sur un couple de carrés qui ne sont pas cochés symétriquement sur les deux volets de la carte, on peut localiser des phrases équivalentes où sont attestées des variations dans la traduction des

³ Le traitement automatique de *corpus* parallèles suppose une phase préalable d'alignement http://tal.univ-paris3.fr/mkAlign/_fn1, c'est-à-dire de mise en correspondance dans chacun des volets de différents types d'unités textuelles (Zimina 2004b). On peut établir des correspondances entre des unités de différents niveaux : mots, syntagmes, phrases, paragraphes, sections, etc.

⁴ Le programme *mkAlign* est développé par Serge Fleury au Centre de lexicométrie et d'analyse automatique des textes (SYLED-CLAT, Université Paris III). Cet outil permet à l'utilisateur de construire un alignement d'un couple de textes et de définir son degré de précision (résolution). Pour plus d'information sur *mkAlign*, on consultera le site Web dédié à cet outil : <http://tal.univ-paris3.fr/mkAlign/mkAlignDOC.htm>

segments étudiés. Dans l'exemple présenté sur la figure 6, ces variations correspondent à des séries de répétitions lexicales légèrement altérées par des modifications de l'un de leurs composants (appelés aussi quasi-segments répétés) : *tout état membre/any [member] state*.

Cet exemple montre que les répétitions des formes graphiques n'ont pas forcément les mêmes origines dans les deux volets du *corpus*. Elles dépendent des contraintes syntaxiques propres à chacune des langues, de la variation lexicale, des choix sémantiques effectués par le traducteur, etc. L'exploration à l'aide de la carte des sections permet d'envisager une analyse approfondie de la hiérarchie de correspondances qui se forment autour des pôles *état/state/государство* dans les deux volets du *corpus* « convention » (cf. tableau 5).

On peut envisager plusieurs types de sélection des unités textuelles pour engager l'exploration à l'aide de la carte des sections parallèles. Par exemple, une « sélection par seuillage » (Zimina 2004a) est amorcée par la sélection dans le texte source des sections dans lesquelles les occurrences d'une unité textuelle (forme, segment répété, cooccurrence, patron morpho-syntaxique) dépassent un seuil fixé. Une « sélection topographique » se réalise à partir de considérations portant sur la localisation des unités bilingues dans les fragments textuels appariés (phrases, paragraphes, sections). Ce type de sélection est particulièrement adapté à l'étude des équivalences lexicales de basse fréquence.

Sur la figure 7, la description cartographique des volets anglais/russe du *corpus* « convention », traduit simultanément la ventilation des types correspondant au segment répété anglais *any state* (F = 14) et le segment russe *любо. + государство*. + (F = 16). Le segment russe est défini sous la forme d'expressions régulières⁵. L'utilisation du langage des expressions régulières permet de tenir compte de désinences qui changent les formes graphiques des mots russes : *любое государство, любого государства*, etc. L'analyse visuelle des deux volets de la carte rend possible une localisation rapide de l'ensemble des contextes où les deux segments bilingues ne sont pas des traductions mutuelles. Pour repérer ces contextes originaux, il suffit de sélectionner des paires de carrés qui ne sont pas cochés symétriquement sur les deux volets de la carte (cf. figure 7) :

Anglais	Russe
<u>any state</u> may/.../ specify the territory	<u>любая сторона</u> /.../ может /.../ указать территорию

Cette exploration montre que la cartographie des « présence-absence » de correspondances bilingues au sein des traductions fournit des moyens automatisés pour le

⁵ Le langage des expressions régulières offre la possibilité de représenter des portions de texte à l'aide d'un ensemble riche d'opérateurs. Il est accessible dans la plupart des systèmes informatiques.

recensement des équivalences, y compris pour des mots qui possèdent plusieurs traductions au sein d'un même *corpus* (Zimina 2004b, 2005).

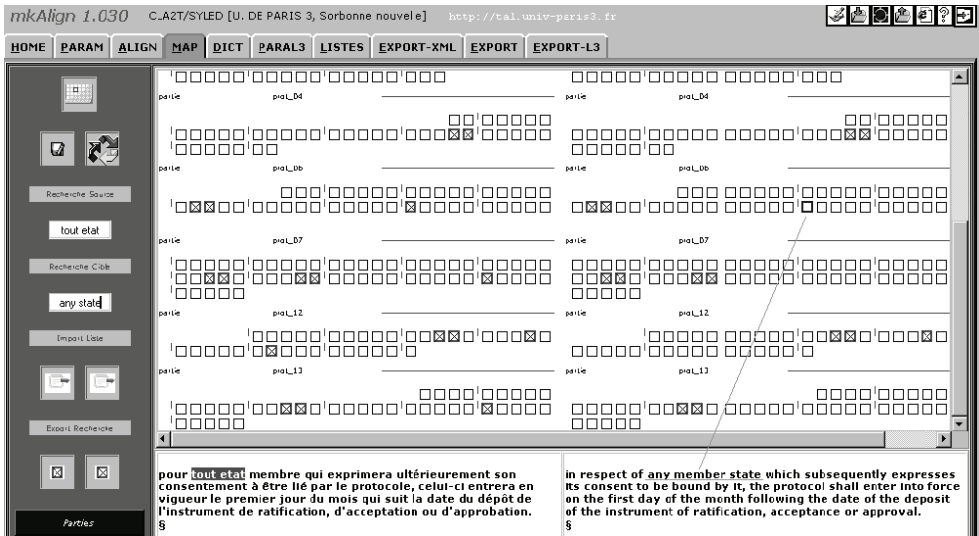


Figure 6 : Cartographie d'équivalences traductionnelles dans les volets français/anglais du *corpus* parallèle « convention »

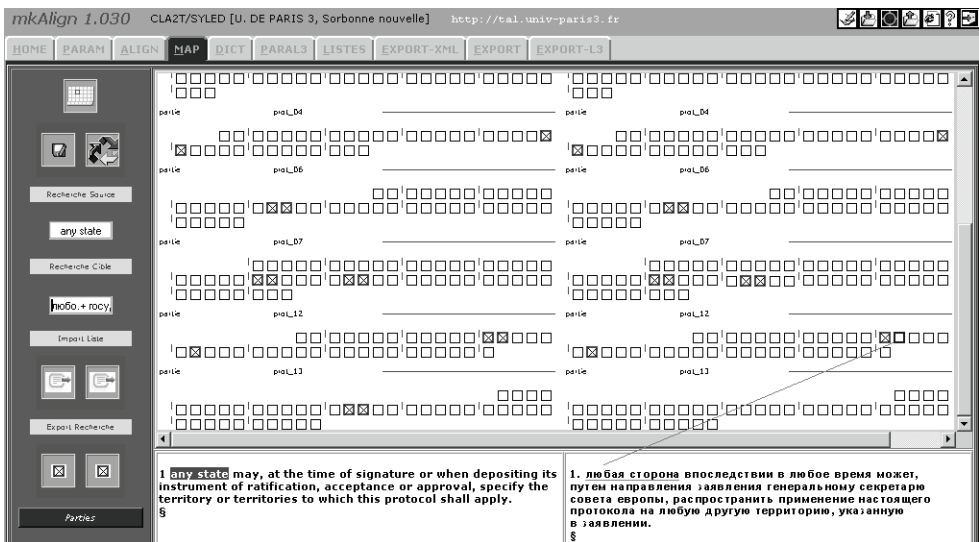


Figure 7 : Cartographie d'équivalences traductionnelles dans les volets anglais/russe du *corpus* parallèle « convention »

4. CONCLUSION

L'analyse textométrique offre une base précieuse pour l'acquisition de ressources traductionnelles à base de *corpus*. La textométrie permet d'accéder à de nouvelles dimensions d'analyse des données de traduction entre des phénomènes langagiers relevant de différents niveaux de l'analyse linguistique.

Au travers de plusieurs explorations du *corpus* « convention » présentées dans cet article, nous avons défini une série de pratiques qui permettent d'accéder à la description de relations de correspondance entre les unités de *corpus* parallèles multilingues. Une approche hybride qui allie la représentation topographique du bitexte et la navigation textométrique en *corpus* est à l'origine de ces explorations.

L'utilisation de la topographie bitextuelle dans le cadre du traitement automatique de *corpus* parallèles multilingues constitue une piste de recherche extrêmement prometteuse. Les applications ouvertes par cette approche concernent notamment la construction de nouvelles procédures informatiques susceptibles de dévoiler la complexité de la structure des équivalences qui se forment au niveau des mots et des syntagmes dans les textes originaux et leurs traductions.

BIBLIOGRAPHIE

- Lamalle (C.), Martinez (W.), Fleury (S.), Salem (A.), Kuncova (A.), Maisondieu (A.) 2003 : *Lexico3 – Outils de statistique textuelle, manuel d'utilisation*, Paris, Université de la Sorbonne nouvelle – Paris III (Travaux du Syled-CLA²T).
- Lamalle (C.), Salem (A.) 2002 : « Types généralisés et topographie textuelle dans l'analyse quantitative des *corpus* textuels » dans Morin (A.) et Sébillot (P.), éd., *Actes des 6^{es} Journées internationales d'analyse statistique des données textuelles*, Saint-Malo, 13-14 mars 2002, INRIA, pp. 403-412.
- Lebart (L.), Salem (A.), 1994 : *Statistique Textuelle*, Paris, Dunod.
- Salem (A.), 1987 : *Pratique des segments répétés : essai de statistique textuelle*, Paris, Klincksieck.
- Zimina (M.), 2004a : « Alignement textométrique des unités lexicales à correspondances multiples dans les *corpus* parallèles » dans Purnelle (G.), Fairon (C.) et Dister (A.), édité, *Le poids des mots*, actes des 7^{es} Journées internationales d'analyse statistique des données textuelles, Louvain-la-Neuve, 10-12 mars 2004, Presses universitaires de Louvain, pp. 1195-1202.
- Zimina (M.), 2004b : « Approches quantitatives de l'extraction de ressources traductionnelles à partir de *corpus* parallèles », thèse de Doctorat en Sciences du langage, soutenue le 26 novembre 2004, Paris, Université de la Sorbonne nouvelle – Paris III.
- Zimina (M.), 2005 (à paraître) : « Bi-text Topography and Quantitative Approaches of Parallel Text Processing » dans *Proceedings from the Corpus Linguistics Conference Series*, vol. I, n° 1, Birmingham, 14-17 juillet 2005, Birmingham University, Centre for *Corpus* Research.

ÉLABORER DES *CORPUS* XML EN LANGUES PARTENAIRES : QUELLES TECHNOLOGIES APPROPRIÉES ? UNE EXPÉRIENCE EN MAURITANIE ET AU SÉNÉGAL

Thierno Cisse

département de linguistique
Université Cheikh Anta Diop de Dakar
Dakar, Sénégal

Paul Muraille – Marc Van Campenhoudt

centre de recherche TERMISTI
Institut supérieur de traducteurs et interprètes
Haute École de Bruxelles
Bruxelles, Belgique
amidiaw@sentoo.sn; paul.muraille@tictips.com; mvc@termisti.be

Résumé : *L'élaboration de corpus textuels informatisés représente un défi majeur pour le développement des langues partenaires. Les technologies Unicode, XML et XSL ont, certes, des potentialités séduisantes, mais leur maîtrise effective peut se révéler délicate face aux réalités du terrain. S'inspirant des notions de technologies et méthodologies appropriées au développement durable, les partenaires de l'action de recherche Expérimentation de normes de balisage en langues partenaires proposent d'envisager une chaîne de traitement qui soit tout à la fois simple, efficace et aisément reproductible. Ils soulignent particulièrement l'enjeu primordial de la graphisation et l'obstacle que constitue la langue anglaise, lingua franca de l'ingénierie linguistique.*

Mots-clés : *corpus, XML, graphisation, langues partenaires, technologies appropriées, Sénégal, Mauritanie*

1. INTRODUCTION

L'action de recherche partagée *Expérimentation de normes de balisage en langues partenaires*¹, financée par le réseau *Lexicologie, Terminologie, Traduction* (LTT) de l'Agence universitaire de la Francophonie (AUF), a eu pour ambition de permettre le balisage XML en écriture Unicode de bases de données textuelles et lexi-

¹ www.termisti.refer.org/ltt/ltt.htm. Partenaires : Université Cheikh Anta Diop de Dakar (centre de linguistique appliquée de Dakar et département de linguistique), Université de Nouakchott (département des langues nationales et de linguistique) et Institut supérieur de traducteurs et interprètes (Haute École de Bruxelles, centre de recherche TERMISTI).

cales. L'idée était « de vérifier l'intérêt des propositions de normes existantes au regard de la réalité de langues souvent négligées par la normalisation internationale » (Cisse *et al.* 2004 : 82) en vue de diffuser sur les inforoutes des données linguistiques propres aux langues du Sénégal et de Mauritanie.

La présente communication revient sur les difficultés rencontrées au cours de cette recherche appliquée, alors que nous ne mesurons pas toutes les réalités du terrain. Le poids et l'influence des travaux des universitaires du Nord étaient bien connus, l'un de nos objectifs étant précisément de montrer qu'ils ne constituaient pas une fatalité. Par contre, nous avons moins bien mesuré la difficulté de récupérer et réunir des *corpus* textuels ou lexicaux déjà informatisés. Plus encore, les contraintes liées aux différentes fractures qui frappent les pays du Sud se sont révélées très prégnantes, au point de nous pousser, au terme de la recherche, à reconsidérer notre approche au travers du prisme des technologies et méthodologies appropriées au développement durable.

2. LES TECHNOLOGIES ET MÉTHODOLOGIES APPROPRIÉES AU DÉVELOPPEMENT DURABLE

2.1. Les technologies appropriées

La notion de technologies appropriées au développement a fait florès dans le monde des ONG au tournant des années soixante-dix à quatre-vingt. Elle est née de l'échec des grands programmes de coopération mis en place dans le sillage de la décolonisation. Le mouvement des technologies appropriées, qui s'est étendu aux questions de méthodologie, s'appuie largement sur l'idée d'un microdéveloppement fondé sur les moyens immédiatement disponibles : « Toute société dispose de technologies qui assurent son développement ou du moins sa survie. Ces technologies sont le résultat de la capacité d'invention et d'adaptation de cette société. Leur degré de sophistication et de complexité varie considérablement d'une société à une autre pour des raisons multiples. » (Crombrugge 1984 : 65.)

Mieux qu'une brève définition, une série non exhaustive de critères permettent d'évaluer l'appropriation d'une technologie aux besoins en développement durable : « l'économie de devises, l'économie d'investissement, l'intensité en main-d'œuvre, l'économie d'énergie, l'utilisation d'énergies renouvelables, la décentralisation de la création de valeur ajoutée, la préservation du potentiel écologique, l'autonomie technique et financière des utilisateurs pour le fonctionnement, l'entretien et la réparation des équipements, la reproductibilité locale, le potentiel de diffusion, l'acceptation par les populations, l'utilisation des matériaux locaux, l'utilisation du savoir-faire et de l'expérience locale. » (Crombrugge 1984 : 65².)

Le mouvement s'est surtout focalisé sur des besoins fondamentaux du développement : l'eau, la terre, l'énergie, la santé. À ce jour, très peu de travaux semblent avoir porté sur la graphisation, la bureautique ou encore la microédition.

² Voir aussi les critères énoncés par Darrow et Saxenian (1993).

2.2. Le développement durable

Soucieuses de l'homme et respectueuses de l'environnement, les techniques et méthodologies appropriées se sont inscrites, avant l'heure, dans la perspective du « développement durable ». Cette dernière notion a été mise en avant en 1987 par les Nations unies au départ du *rapport Brundtland* (CMED 1988). En 1992, la Conférence des Nations unies sur l'environnement et le développement (CNUED) a adopté à Rio vingt-sept principes devant permettre un tel développement et établi un guide de mise en œuvre du développement durable au XIX^e siècle, le fameux *Agenda 21*.

2.2.1. Le rôle de la langue commune et des langues partenaires

Le rôle primordial des langues partenaires n'a guère été souligné dans le cadre de cette dynamique et la question linguistique n'est guère abordée de manière spécifique dans les grands textes de référence³. La Francophonie a toutefois le mérite particulier d'avoir intégré la question de la diversité culturelle et linguistique lors du récent colloque *Développement durable : leçons et perspectives* (Ouagadougou, 1^{er} au 4 juin 2004). On est particulièrement heureux de lire dans le résumé des principales recommandations que « les participants au colloque réaffirment le caractère inaliénable de la diversité culturelle et linguistique comme fondement du développement durable⁴ ».

2.2.2. Développement durable et fracture numérique

Les débats concernant les différentes fractures – médicale, éducative, sociale, numérique... – qui affectent le développement posent, certes, la question de l'adéquation des TIC aux besoins réels (*cf.* Guichard 2003 et Valérien et Wallet 2004 : 119). On conviendra toutefois que l'ingénierie linguistique offre aujourd'hui des possibilités extraordinaires de description des langues partenaires, souvent en danger, et permet d'envisager une communication efficace entre locuteurs de ces langues. Le lien entre le développement des langues partenaires et celui des technologies de l'information apparaîtra évident à la communauté des linguistes, alors qu'il n'a pas toujours été explicité, loin s'en faut, dans les grandes déclarations officielles⁵.

2.3. Ingénierie linguistique et technologies appropriées

Notre expérience nous amène à souligner combien les conditions de travail des chercheurs du Sud compliquent l'émergence d'une ingénierie linguistique autonome. Il nous semble donc primordial de définir les technologies et méthodologies les plus adéquates pour leur permettre de contribuer au mieux à l'étude et au développement de leurs langues, conçues comme les médiums appropriés d'un déve-

³ Le *Glossaire pour le développement durable*, qui propose une traduction en français des termes les plus utilisés ne contient quasiment aucun terme lié aux sciences du langage (www.francophonie-durable.org/textes.html).

⁴ www.francophonie-durable.org/documents/Principales_recommandations.pdf.

⁵ Voir, par exemple, le plan d'action issu de la 3^e *conférence ministérielle sur la culture* organisée par l'AIF à Cotonou les 14 et 15 juin 2001.

loppement durable. Nous n'entendons pas évoquer ici la nécessité d'un matériel informatique approprié, exploitant des logiciels libres et dont l'énergie serait fournie par une manivelle⁶. Il s'agit plus modestement d'envisager la méthodologie logicielle la plus appropriée à notre objectif initial de création de *corpus* linguistiques au Sénégal et en Mauritanie. Notre réflexion nous conduit ainsi à envisager divers critères d'évaluation des voies les plus appropriées à cet objectif. Au regard de celui-ci et des situations locales, tous ces critères n'auront bien sûr pas le même poids mais peuvent constituer autant de repères au moment d'opter pour la mise en œuvre d'une chaîne de traitement de *corpus*⁷. Certains sont d'ordres technique et financier : gratuité ou faible coût des logiciels ; installation depuis un support « ancien » (disquette, CD-Rom) ; Internet non indispensable ; fonctionnement multiplateforme, même sur un système daté ; faible consommation de ressources système (mémoire vive, espace disque, etc.). D'autres concernent les fonctionnalités mêmes des logiciels : compatibilité avec la norme Unicode ; entrées et sorties en formats non-proprétaires et, dans la mesure du possible, stockage interne dans un format ouvert (p. ex. XML) ; modularité et intégration des fonctionnalités (le même logiciel traite plusieurs étapes).

L'aspect humain doit également être considéré. Il importe de garantir l'autonomie d'un chercheur qui ne peut ni recourir à un informaticien ni s'improviser informaticien. Il convient donc d'éviter, d'une part, qu'il doive effectuer de nombreuses manipulations complexes de données au travers d'une succession de logiciels distincts et, d'autre part, que la prise en main de chaque outil se révèle ardue. À cet égard, un critère essentiel de l'appropriation est la disponibilité de l'interface et de la documentation (hors ligne) des outils dans la langue commune (le français)⁸.

3. ACCÉDER À L'INFORMATION EN LANGUE FRANÇAISE

Les ONG actives dans le domaine des technologies appropriées se sont toujours souciées de la communication des savoirs pratiques, au travers de centres de documentation ou sous la forme de fiches simplifiées, rédigées dans un français accessible. Certaines, comme l'INADES (Institut africain pour le développement économique et social), ont particulièrement été attentives à l'utilisation du français fondamental. Des manuels ont parfois même été rédigés en langues partenaires (p. ex. Schmitz 1986).

⁶ Pour ce type de problématique, voir le *Hundred-Dollar Laptop Project* Media Lab (MIT) : <http://laptop.media.mit.edu>.

⁷ Nous assumons le fait que certains de ces critères sont incompatibles : l'émergence d'Unicode est liée au progrès informatique et suppose un système d'exploitation récent. Des passerelles avec des systèmes plus anciens sont heureusement possibles (Chanard et Popescu 2001).

⁸ Ce critère est une condition nécessaire mais non suffisante d'appropriation : quand bien même une interface et une documentation sont disponibles en langue commune, il convient ensuite d'en vérifier la qualité sous l'angle de la forme (qualité linguistique) et sous celui du fond (richesse documentaire, p. ex. disponibilité d'un didacticiel, de fichiers d'exemples, d'un manuel de l'utilisateur et d'un manuel de référence, etc.).

3.1. Une *lingua franca* incontournable

L'anglais est la langue véhiculaire de l'informatique. La très grande majorité des documents indispensables pour s'appropriier les technologies de l'ingénierie linguistique ne sont disponibles qu'en anglais⁹, alors même qu'ils ont parfois été rédigés par des francophones. La seule manière de participer au débat scientifique – par exemple pour faire évoluer la norme Unicode (cf. 4) – est de s'exprimer dans cette langue. Force est, malheureusement, de constater que celle-ci demeure très mal maîtrisée par nombre de diplômés du Sud éduqués dans l'idée de la grandeur de la langue métropolitaine. Ne pouvant lire des documents indispensables, ils se trouvent très rapidement bloqués dans leurs possibilités d'élaborer des méthodologies fondées sur le génie logiciel. Un rapide inventaire des normes et programmes disponibles prouve vite qu'il est impossible de travailler sans bien comprendre l'anglais. Des traductions fragmentaires existent parfois, mais elles sont le plus souvent le fruit d'initiatives dispersées et le fait de bénévoles non formés à la traduction. En outre, les mises à jour des documents originaux sont rarement prises en compte.

3.2. Pour une cellule de traduction « inforoutes »

Il serait vain, sinon puéril, de se battre pour obtenir que le français devienne la langue internationale du progrès informatique. Il est, bien sûr, attristant de constater que des chercheurs francophones ne diffusent pas également dans leur propre langue les normes internationales qu'ils ont rédigées, nécessité fait loi, en anglais. Il serait toutefois injuste de leur faire le procès de ne pas se soucier de l'existence d'une version en langue française : la traduction exige, outre une formation particulière, de disposer de moyens logistiques et financiers adéquats. Imposer le français sur le théâtre de la conception des autoroutes de l'information est un combat déjà perdu. Assurer une traduction professionnelle rapide permettrait, par contre, d'éviter qu'un fossé encore plus important ne se creuse entre le Sud et le Nord.

Les linguistes francophones n'occuperaient-ils pas davantage le terrain de la normalisation des échanges de données s'ils pouvaient disposer très rapidement d'une traduction officielle de toutes les publications importantes ? La Francophonie institutionnelle serait assurément bien inspirée de financer un programme de traduction de l'anglais de tous les documents nécessaires à une appropriation rapide des savoirs indispensables à la création des inforoutes en français et en langues partenaires¹⁰. Traduire ensuite vers l'anglais les réactions des linguistes du Sud permettrait d'interagir rapidement avec les créateurs de normes pour veiller à la prise en compte des besoins propres aux langues partenaires. Par ailleurs, la localisation des quelques logiciels incontournables pour la description linguistique faciliterait grandement le travail de nombreux chercheurs qui n'arrivent guère à en exploiter toutes

⁹ Norme Unicode, normes du W3C, normes d'échange de données linguistiques, logiciels de balisage, concordanciers...

¹⁰ Une retombée indirecte mais non négligeable de l'activité d'une telle cellule pourrait être la standardisation de la terminologie de ce sous-domaine des systèmes d'information, en coopération avec les organismes francophones compétents.

les potentialités¹¹. Que coûterait le financement annuel d'une cellule légère de deux ou trois traducteurs délocalisée au Sud au regard du coût de l'organisation d'événements qui tiennent plus de la vitrine éphémère que du développement durable ?

4. LA PROBLÉMATIQUE DE LA GRAPHISATION

La graphisation des langues partenaires constitue un enjeu majeur. Elle est la condition *sine qua non* de leur traitement informatique et donc de leur utilisation sur le réseau Internet. Certains proposent de simplifier leur écriture pour éviter les difficultés de numérisation générées par les diacritiques et autres signes moins fréquents (voir, par exemple, Chaudenson 2004). On devine aisément que les caractères empruntés à l'alphabet phonétique international sont les premiers visés par les tenants de ce point de vue.

Cette proposition semble, *a priori*, inspirée par la philosophie des technologies appropriées. Elle néglige malheureusement la question fondamentale de l'appropriation : on ne peut nier aux locuteurs des langues partenaires le droit de choisir, eux aussi, la meilleure manière d'écrire leur langue. Les peuples européens ont-ils abandonné leurs alphabets pour se conformer à la norme américaine ASCII fondée sur les 26 lettres de l'alphabet latin ? La loi du marché et les progrès technologiques ont plutôt débouché sur une adaptation des méthodes de codage aux besoins des consommateurs : la norme Unicode, censée inclure toutes les écritures du monde dans une table unique, sans nécessité de procéder à des transcodages.

4.1. Les réalités du terrain

Au Sénégal et dans les pays de la même aire linguistique, il existe une longue tradition de description des traditions orales et de constitution de *corpus* oraux. Les textes législatifs sur l'écriture des langues nationales sénégalaises datent des années 70 et depuis, il n'y a jamais eu de politique ardue de développement des langues locales et d'incitation à l'écriture standardisée. Le français demeure la langue officielle et c'est seulement depuis une décennie, avec le développement des radios libres, que les langues locales, et notamment le wolof, ont commencé à concurrencer le français, mais uniquement au niveau discursif.

Les premières transcriptions n'ont pas toujours tenu compte des besoins de constitution de *corpus* écrits d'envergure et de partage des données ; les alphabets utilisés sont différents d'un chercheur à l'autre. Une graphisation harmonisée à l'échelle des aires linguistiques qui permette une communication scientifique et une présence sur Internet demeurent indispensables pour la survie des langues et cultures locales. Il s'agit de reprendre systématiquement les transcriptions existantes et en même temps de vulgariser les outils modernes de prise en charge des langues locales.

¹¹ Nous pensons particulièrement au célèbre logiciel *Toolbox* (ex-*Shoebox*).

La solution de la graphisation harmonisée, si elle veut être opérante, doit tenir compte des problèmes logistiques : l'électricité déficiente, l'équipement informatique réduit, souvent vieillissant et fort différencié d'un laboratoire à l'autre dans le même pays et d'un pays à l'autre, les antivirus inefficaces, les connexions lentes – lorsqu'elles existent – et la pénurie de personnel informatique affecté aux recherches en sciences humaines et sociales. Cette solution tiendra ainsi compte de la fameuse « fracture numérique » dans le choix des logiciels et des technologies les plus appropriées, respectueuses de l'expression locale.

4.2. Résoudre la question du clavier, une étape indispensable

L'absence de véritable marché explique vraisemblablement la non-commercialisation de claviers d'ordinateur propres aux grandes langues africaines. La faiblesse économique explique sans doute également que les écritures des langues d'Afrique de l'Ouest ne sont pas encore intégrées comme telles dans Unicode. L'indigence pécuniaire n'est cependant pas nécessairement un obstacle : on devine que seul un intense travail de *lobbying* a permis l'intégration dans Unicode des hiéroglyphes, des syllabaires autochtones canadiens, du syllabaire éthiopien et, depuis peu, de l'écriture tfinagh du berbère. Une récente démarche, soutenue par le monde anglophone, est en passe de faire aboutir l'intégration de l'écriture n'ko du mandingue dans Unicode, preuve de la nécessité de soutenir des initiatives volontaristes. La Francophonie pourrait aisément financer la participation régulière d'experts africains aux activités du consortium Unicode. Ceci supposerait, encore une fois, que soit créée une cellule de traduction apte à soutenir les chercheurs travaillant sur la graphisation des langues partenaires (*cf.* 3.2), à moins de préférer qu'ils ne s'associent aux universités d'outre-Atlantique, très attentives à la question¹².

4.3. Le clavier virtuel, solution simple et bon marché

La possibilité de réaffecter les touches d'un clavier physique est une procédure alléchante. Dans le cadre de notre projet, deux logiciels ont été expérimentés : *Keyman Developer* et *Microsoft Keyboard Layout Creator (MKLC)*¹³. Comme les partenaires du projet l'ont déjà montré (Cisse *et al.* 2004 : 86-92), les claviers créés¹⁴ fonctionnent dans un grand nombre d'applications courantes, offrent la certitude d'utiliser le bon caractère Unicode, sont gratuits, tiennent sur une disquette et s'installent aisément. Il n'est pas nécessaire de connaître Unicode pour les utiliser au quotidien, mais leurs interfaces anglaises devraient être localisées en français.

La création de ce type de clavier suppose d'identifier les caractères adéquats dans Unicode, en l'absence d'un véritable « bloc » qui soit consacré aux langues concernées. Le petit gratuiciel *Babelmap*, dont l'archive pèse à peine 600 Ko et qui

¹² Voir, par exemple, la très intéressante *Script Encoding Initiative* de l'Université de Californie (Berkeley) : www.linguistics.berkeley.edu/sei/index.html.

¹³ www.tavultesoft.com/keymandev et www.microsoft.com/globaldev/tools/msklc.msp.

¹⁴ Téléchargeables à l'adresse : www.termisti.refer.org/lft/lft03.htm.

ne nécessite aucune connexion à Internet, constitue sans doute une solution simple et maniable pour qui souhaite s'atteler à cette tâche. Disponible en français, il propose également une documentation Unicode dans cette langue¹⁵.

Le processus de création de claviers proprement dit ne devrait intéresser *a priori* que des chercheurs en linguistique. Le gratuiciel *MKLC* s'est révélé plus simple que *Keyman Developer*, même si les potentialités de ce dernier sont très intéressantes pour créer un clavier très étendu. Malheureusement, ces deux interfaces et leurs documentations n'existent qu'en anglais.

Les claviers créés à l'aide de *Keyman Developer* ou de *MKLC* sont relativement simples d'emploi, mais les premiers se révèlent parfois capricieux. Les claviers *MKLC* sont intégrés à *Windows* : ils peuvent donc s'installer dans une interface française et s'utiliser selon la classique procédure de basculement de clavier propre à ce système d'exploitation. Il importe toutefois d'observer que si *Keyman* et *MKLC* fonctionnent avec les programmes de bureautique les plus courants, ainsi qu'avec des éditeurs XML comme *XML Spy* et *Oxygen*, certains logiciels bien connus ne les acceptent pas nécessairement¹⁶.

5. QUELLE CHAÎNE DE TRAITEMENT DU CORPUS TEXTUEL ?

Notre objectif fondamental était la mise en commun de données linguistiques en langues partenaires grâce à l'usage conjoint d'Unicode et d'une norme XML. Nous pensions que le mécanisme de transformation des documents XML à l'aide du langage de feuilles de style XSL permettrait, dans un second temps, une diffusion aisée des textes sur la toile. Dans ce cadre, les standards d'échange XCES (*Corpus Encoding Standard for XML*) et TEI (*Text Encoding Initiative*), qui permettent d'enrichir le document par une grande variété de descripteurs, ont été principalement testés.

5.1. L'utilisabilité des normes XML

L'échange de données lexicales selon le standard XML semble relativement aisé à maîtriser dès lors que le modèle de données a été bien pensé et que les catégories de données sont restreintes¹⁷. La maîtrise des normes textuelles XCES et TEI s'avère sans doute plus délicate au vu de la masse d'informations à appréhender. La concurrence entre les normes TEI et XCES est regrettable dans la mesure où leurs différences ne semblent pas évidentes au premier abord. L'impression de solitude est grande pour qui souhaite s'initier en autodidacte à ces normes et arriver à une bonne maîtrise. Quel que soit le contenu à baliser, le principal écueil réside, encore une fois, dans la nécessité de bien comprendre la langue anglaise. Seule la

¹⁵ www.babelstone.co.uk/Software/BabelMap.html.

¹⁶ Tel est le cas du célèbre éditeur HTML *Web Expert 6*. Les claviers *MKLC* ne semblent pas fonctionner dans le logiciel *Toolbox*, lequel fonctionne très bien avec *Keyman*.

¹⁷ Le traitement des données lexicales et terminologiques, également abordé dans le cadre de la recherche, n'est pas envisagé ici.

TEI *Lite* possède une traduction officielle, particulièrement bienvenue¹⁸ : le souhait d'aller à l'essentiel en fait un bon outil en termes de technologies appropriées, même si l'on doit regretter que la simplification ait réduit la portée de la norme aux textes « littéraires », sans véritable prise en compte des catégories propres à la littérature orale. Par ailleurs, la volonté des concepteurs de la TEI de proposer une interface capable de générer une « sous-DTD » adaptée à des besoins spécifiques¹⁹ est particulièrement louable.

5.2. Éditeurs XML spécifiques

On trouve aisément sur la toile des éditeurs XML gratuits ou à coût réduit. Outre que beaucoup n'offrent que peu de fonctionnalités et souffrent de l'absence d'une documentation exhaustive, la plupart ne sont, ici encore, disponibles qu'en anglais. L'action de recherche ne visant pas à les expérimenter spécifiquement, les partenaires ont travaillé successivement avec deux produits renommés et de coût abordable : *XML Spy* et *Oxygen*²⁰. Le deuxième présente le double avantage d'être localisé et documenté en français et de proposer des feuilles de style adaptées au traitement de documents XML conforme à la TEI.

Le public cible de ces logiciels est relativement restreint. Les linguistes qui ne sont pas informaticiens mais connaissent XML n'y auront vraisemblablement recours que pour disposer de facilités d'encodage, de validation du document et de conversion à l'aide de feuilles de style. Les points essentiels pour le choix d'une interface seraient donc logiquement : être disponible en français, répondre aux critères de choix des logiciels en termes de technologies appropriées, faciliter le travail de balisage, permettre d'associer aisément une DTD et une feuille de style, posséder un moteur de transformation garantissant le rendu des caractères Unicode propres aux langues traitées²¹.

Par ailleurs, le langage de transformation XSL s'avère rapidement complexe, ce qui constitue un obstacle majeur au regard du critère d'autonomie des chercheurs du Sud. Il importe donc que soient rendues disponibles davantage de feuilles de style permettant de valoriser les *corpus* balisés. Malheureusement, les concepteurs des normes XCES et TEI ne semblent guère concernés par cette tâche, qui échappe à leur activité de normalisation.

5.3. Traitement de texte proposant des portes d'entrée et de sortie vers XML

L'expérience nous donne à penser qu'un éditeur XML n'est pas un outil suffisamment simple pour nombre de linguistes qui engrangent des textes sans souhaiter nécessairement exploiter un balisage XML fin. Plutôt que de leur demander un

¹⁸ La TEI simplifiée : www.tei-c.org/Lite/teiu5_fr.html.

¹⁹ www.tei-c.org/pizza.html.

²⁰ Hormis les deux éditeurs commerciaux susmentionnés, divers logiciels à code ouvert (*Bitflux*, *Jaxe*, *jEdit*, *Peter's XML Editor*, *TreeBeard* et *Vex*) ont été mis à l'essai mais ont été écartés car ils ne satisfaisaient pas à un ou à plusieurs des critères d'appropriation évoqués.

²¹ Le moteur de transformation propre à la version 2004 de *XML Spy Home Edition* s'est avéré décevant de ce dernier point de vue.

balisage lourd et fastidieux, il semble préférable de leur proposer d'encoder leur *corpus* dans un traitement de texte selon une logique de document structuré : utiliser des styles renvoyant aux catégories de données *ad hoc* d'une norme d'échanges devrait permettre ensuite une conversion vers un format XML. L'expérience nous suggère que cette solution satisferait nombre de linguistes ou d'étudiants en sciences du langage qui ne peuvent s'investir dans des technologies de l'information trop complexes.

Le consortium de la *Text Encoding Initiative* propose déjà une semblable procédure pour le traitement de texte *Writer* d'*OpenOffice*. Elle permet de sauvegarder un fichier *Writer* au format XML de la TEI et, inversement, d'ouvrir un fichier XML dans *Writer*. L'expérimentation montre la viabilité de cette solution qui présente le triple avantage d'être gratuite, aisée et peu gourmande en ressources informatiques. Certes, les fichiers proposés (DTD et XSL) ne dépassent pas la simple démonstration. Cependant, il semble tout à fait possible pour un spécialiste de les améliorer en veillant à diversifier les éléments présents dans la DTD et les styles qui leur correspondent dans le modèle de document proposé²².

La plateforme *Cyberdocs* du projet *Cyberthèses*²³, qui publie sur la toile des thèses du monde entier, exploite une approche similaire. Développée avec l'aide de l'Agence de la Francophonie, elle permet, en effet, de convertir vers la *TEI Lite*, et via *OpenOffice*, des fichiers de traitements de texte. Nous n'avons malheureusement pas pu vérifier si une telle solution garantit bel et bien un rendu correct des caractères Unicode (Abrahamane 2004). À l'instar de *Cyberthèses*, l'Agence bibliographique de l'enseignement supérieur a mis au point un système de publication de thèses en ligne basé sur la TEI (*Sparte*²⁴) ; à la différence toutefois que le format supporté avant la transformation XML est le format RTF. Divers autres projets francophones ne mettent pas en œuvre un balisage TEI avec *OpenOffice*, mais fournissent des modèles d'implémentation du couple *OpenOffice/XML* reposant sur les principes exposés ici et démontrent la faisabilité de la démarche proposée. Citons à titre de référence *Ooo2Dbk*, une chaîne de production documentaire mise en œuvre au sein du ministère français de l'Équipement²⁵ ou la plateforme de publication en ligne *Lodel*²⁶.

6. EN GUISE DE SYNTHÈSE : QUELQUES MESURES PRATIQUES

6.1. La mise en ligne de fiches pratiques

Tout linguiste confronté à des problèmes informatiques a un jour éprouvé la difficulté de trouver une information cohérente, simple, pratique et rédigée en français sans devoir parcourir un grand nombre de sites aux réponses fragmentaires. Le mouvement des technologies appropriées a souvent veillé à transmettre ses connaissances à travers des fiches

²² Cette remarque concerne particulièrement les passages versifiés et dialogués.

²³ sourcesup.cru.fr/cybertheses.

²⁴ www.abes.fr/abes/DesktopDefault.aspx?tabid=315.

²⁵ www.indesko.com/fr/telechargements/ooo2dbk.

²⁶ www.lodel.org.

pratiques (on songe aux célèbres fiches du GRET²⁷) qui ont parfois été informatisées (p. ex. Agridoc²⁸). En ingénierie linguistique, de telles fiches, rédigées en un français accessible, permettraient à nombre de chercheurs, d'enseignants et d'étudiants du Sud comme du Nord de se débrouiller au quotidien pour implémenter des logiciels simples, disposer des polices Unicode adéquates, configurer correctement leur navigateur ou leur logiciel de courrier, produire un document XML, opérer une transformation XSL, etc.

Il serait également utile d'associer à de telles fiches pratiques des exemples concrets d'application aisément adaptables sur le terrain. On songera, par exemple, à des descriptifs Unicode des principaux caractères utilisés par les africanistes, à des claviers virtuels ou encore à des modèles d'application XSLT.

6.2. Des ressources francophones plus nombreuses et aisément disponibles

La barrière de la langue anglaise plaide pour la mise en place d'un répertoire commenté des ressources francophones déjà disponibles en matière d'ingénierie linguistique. Une série, certes restreinte, de ressources francophones de qualité existent bel et bien sur la toile. Toutefois, leur visibilité est extrêmement faible parce qu'elles sont dispersées, résultent d'initiatives personnelles ou ne traitent qu'incidemment de contenus spécifiques à l'ingénierie linguistique. D'un coût anecdotique, une telle initiative permettrait de centraliser les informations indispensables à la valorisation des langues partenaires.

Une politique volontariste de localisation des logiciels incontournables et de traduction des normes fondamentales disponibles uniquement en anglais – et de leurs mises à jour – permettrait de rapidement multiplier de semblables ressources. Son financement suppose un changement de stratégie dans la défense de notre langue commune : passer de la défense du pré carré francophone à un dialogue constructif et intelligent avec les locuteurs de la nouvelle *lingua franca* de la connaissance.

6.3. La diffusion des technologies Unicode

Le clavier d'ordinateur constitue le point d'entrée du contenu numérique. Il ne coûterait guère de veiller à ce que les locuteurs des langues partenaires puissent écrire aisément leur langue à l'aide de l'outil informatique. Sans devoir miser sur des programmes prestigieux, il suffirait de soutenir des projets de diffusion de ces petits claviers virtuels Unicode auprès des milliers d'utilisateurs potentiels : écoles, centres de santé, municipalités, ministères, presse, gestionnaires de site Internet, ONG... Ils sont, en effet, une condition indispensable à la graphisation des langues, donc à leur utilisation sur un support informatique et sur les inforoutes. Aider à configurer en Unicode les logiciels les plus courants serait une mesure de soutien également aisée à mettre en œuvre.

²⁷ Groupe de recherche et d'échange technologiques (www.gret.org).

²⁸ www.agridoc.com/fichestechniques_gret/index.htm.

6.4. La promotion du document structuré

La maîtrise de XML n'est pas aisée pour qui souhaite créer des *corpus* et elle suppose une pratique régulière. Trop peu d'outils simples sont mis à la disposition des candidats utilisateurs et autres néophytes : les promoteurs des échanges structurés semblent préférer s'arrêter après une démonstration de faisabilité. Une utilisation quotidienne de ces technologies n'est donc viable que pour des institutions possédant une équipe d'informaticiens aptes à développer des interfaces simplifiées.

À défaut de pouvoir convertir la communauté des linguistes aux logiciels de la galaxie *LaTeX*, qui produisent des documents structurés, il paraît plus sage de proposer des formations à une utilisation rationnelle des modèles de document de logiciels comme *Word* ou *Writer*. En effet, l'association d'un style à une catégorie de données permettant d'envisager une conversion vers un document XML (cf. 5.3), il semble plus réaliste de mieux segmenter les compétences : la tâche du linguiste s'arrêterait à l'engrangement selon un modèle de document conforme à une norme d'échange, celle du spécialiste XML étant de faciliter la conversion vers des standards internationaux et de créer des feuilles de style XSL immédiatement utilisables.

BIBLIOGRAPHIE

- Abrahamane (A.), 2004 : « Cyberthèses : une solution à la visibilité de la science africaine ? », dans *Actes de la conférence sur la publication et la diffusion électronique, Dakar, Sénégal, 1-2 septembre 2004*, Dakar : Codesria, www.codesria.org/Links/conferences/el_publ/Abrahamane_Anne.pdf.
- Chanard (Chr.) et Popescu-Belis (A.), 2001 : « Encodage informatique multilingue : application au contexte du Niger », dans *Cahiers du Rifal*, décembre 2001, n° 22, pp. 33-45.
- Chaudenson (R.), 2004 : « La graphisation des langues africaines », dans *Cahiers du Rifal*, décembre 2005, n° 24, pp. 66-67.
- Cisse (Th.), Mbodj (Ch.), Van Campenhoudt (M.) et Wane (M.), 2004 : « Expérimentation de normes de balisage en langues partenaires », dans *Actes des premières Journées scientifiques communes des réseaux de chercheurs concernant la langue « Penser la Francophonie, concepts, actions et outils linguistiques »*, Université de Ouagadougou, 31 mai-1^{er} juin 2004, pp. 81-93, consultable à l'adresse www.bibliotheque.refer.org/livre244/124410.pdf.
- Commission mondiale sur l'environnement et le développement (Cmed), 1988 : *Notre avenir à tous*, Montréal, Éditions du Fleuve – Les publications du Québec.
- Crombrugge (G. de), 1984 : « Favoriser le développement et l'appropriation de la technologie », dans *Le Courrier*, janvier-février 1984, n° 83, pp. 65-66.
- Darrow (K.) et Saxenian (M.), 1993 : *Appropriate Technology Sourcebook*, Stanford : Volunteers in Asia. Consulté à l'adresse : villageearth.org/atnetwork/atsourcebook.
- Guichard (É.), 2003 : « Does the "Digital Divide" Exist ? », dans Seters (P. van), Gaay Fortman (B. de) et Ruijter (A.), dir., *Globalization and its new divides : malcontents, recipes, and reform*, Amsterdam : Dutch University Press. Traduction française : barthes.ens.fr/atelier/geo/Tilburg.html.
- Schmitz (J.-L.), 1986 : *Mutwisi ti bangombe na yandi. Bisalu ya kutwila bangombe na babwala na ndambu ya westi ya Zaïre*, Kinshasa : Inades-Formation, 221 pp. (publié en français sous le titre *L'éleveur et son bétail. L'élevage bovin villageois dans l'ouest du Zaïre*).
- Valérien (J.) et Wallet (J.), 2004 : « À quelles conditions un projet intégrant les TIC dans l'éducation peut-il être considéré comme au service du développement durable ? », dans *Actes du colloque « Développement durable : leçons et perspectives » (Ouagadougou 1^{er} au 4 juin 2004)*, pp. 117-122. Consulté à l'adresse www.francophonie-durable.org/sommaire.html.

UNE COMPOSITION DE SOLUTIONS GÉNÉRIQUES POUR CRÉER LE DICTIONNAIRE FEV ET IMPORTER LE VIETNAMIEN DANS PAPILLON

Vo-Trung Hung¹, Phan-Huy Khanh¹, Christian Boitet²

¹Université de Danang, 54 Nguyen Luong Bang, Danang, Viêt Nam

²GETA-CLIPS-IMAG, B. P. 53, 38041 Grenoble CEDEX 9, France

khanhph@vnn.vn; Hung.Vo-Trung@ud.edu.vn; Christian.Boitet@imag.fr

Résumé : Cet article présente l'application d'une méthode générique de construction semi-automatique de dictionnaires multilingues à la création du FEV, un dictionnaire d'usage français-vietnamien de 20 000 entrées, 50 000 acceptions, 10 000 exemples et 8 000 tournures par ajout automatisé et vérification linguistique des équivalents vietnamiens dans le dictionnaire français-anglais-malais FEM déjà disponible. Le résultat de ce travail est un « premier jet » complet, disponible sur le site web du projet Papillon et dans diverses applications. Une étape de révision « collaborative » sur le web permettra d'améliorer la qualité pour arriver à celle du FEM. L'étape suivante consistera en une recherche collaborative importante menant à la découverte et à la collecte de multiples informations très détaillées concernant le vietnamien, à leur intégration au dictionnaire « DiCo » de Papillon-Nadia, et à la possibilité d'inverser le sens de traduction et de créer des dictionnaires de qualité entre le vietnamien et les autres langues traitées de façon comparable.

Mots-clés : dictionnaire d'usage, français, vietnamien, anglais, malais, FEV, FEM, ingénierie lexicographique, applications dictionnairiques, bases de données lexicales multilingues, projet Papillon

1. INTRODUCTION

Le traitement informatique du vietnamien a commencé dans les années quatre-vingt. Parmi les domaines en technologies de l'information en fort développement dans ce pays, on note la construction de dictionnaires monolingues, bilingues ou trilingues. Cependant, il manque toujours des dictionnaires bilingues d'usage entre français et vietnamien, utilisables aussi bien par les vietnamophones que par les francophones. De plus, les ouvrages disponibles sur papier n'existent pas sous des formes informatiques utilisables aisément sur micro, et ne sont pas accessibles par Internet.

Le projet FEV (réalisation d'un dictionnaire d'usage informatisé français-vietnamien via l'anglais) a été partiellement soutenu par l'AUF (réseau *Francil*). Il a été mené en coopération par l'ERTA (équipe de recherche en traduction automatique, faculté d'informatique, École polytechnique, Université de Danang, Viêt Nam) et le GETA (Groupe d'étude pour la traduction automatique, laboratoire CLIPS, IMAG, UJF et CNRS, Grenoble, France) depuis 1999-2000.

Ce projet a utilisé comme point de départ le dictionnaire FEM (*French-English-Malay*, français-anglais-malais) qui a été réalisé entre 1993 et 1996, puis amélioré 2 fois entre 1998 et 2003 dans le cadre d'une coopération entre le GETA, l'UTMK (*University of Sains Malaysia*, Penang), le DBP (*Dewan Bahasa dan Pustaka*, Kuala Lumpur) et l'Ambassade de France à Kuala Lumpur. La version informatique de ces résultats a été diffusée sous plusieurs formes (disquettes, CD-ROM en HTML...), et est aussi disponible sur le serveur du projet *Papillon* (<http://www.papillon-dictionary.org/>). Le dictionnaire FEM contient environ 20 000 entrées, 50 000 acceptions, 10 000 exemples et 8 000 tournures idiomatiques.

Dans les trois sections qui suivent, nous présenterons plus précisément la structure informatique et linguistique du dictionnaire FEM pris comme point de départ, puis les problèmes particuliers au traitement du vietnamien, la méthodologie générique que nous avons développée pour introduire le vietnamien dans le FEM et produire le FEV, et enfin les applications que nous avons déjà réalisées à partir du FEV.

Du point de vue quantitatif, l'état actuel du FEV correspond au FEM. Du point de vue qualitatif, il n'est pas encore au même niveau. Il s'agit d'une première version, rendue disponible sous plusieurs formes, et il reste à effectuer une révision approfondie par des personnes compétentes en lexicographie contrastive du français et du vietnamien. Notre prochain projet est de réaliser une interface de révision disponible sur le web, si possible intégrée au site *Papillon*, et offrant une ergonomie comparable à celle de la méthode antérieure (édition de fichiers *Word* « stylés »). Les perspectives à plus long terme seront détaillées dans la conclusion.

2. POINT DE DÉPART DU PROJET FEV : LE DICTIONNAIRE FEM

La forme logique du dictionnaire FEM dont nous sommes partis est un ensemble d'articles divisés en acceptions repérées par des étiquettes prises dans une liste fermée (domaines, registre, région) ou des gloses libres. L'anglais a été utilisé en 1993-1994 comme « point de passage » pour construire, par croisement de dictionnaire et lexiques français-anglais avec un dictionnaire anglais-malais, une toute première version. Dès 1995, la structure des articles est devenue « furcoïde », ou « en éventail », c'est-à-dire qu'une acception française mène d'une part à un ou des équivalents anglais, et d'autre part à un ou des équivalents malais. Environ 12,5 % du FEM ont été complétés par des équivalents thaïs en 1996-1997, ces équivalents étant insérés « à côté » des équivalents anglais et malais.

De premiers essais avec une base de données 4D avaient prouvé que le travail lexicographique s'accommode très mal des contraintes d'un SGBD (vue parcellaire, exigence de cohérence à chaque étape), et que ce qu'il faut aux lexicographes, ce sont des éditeurs « syntaxiques », permettant de traiter le dictionnaire comme un document, de faire des copier-coller librement, tout en garantissant la cohérence (formelle) au moins au moment d'une sauvegarde. Or aucun système de création d'un éditeur syntaxique adapté à l'édition d'un dictionnaire multicible n'était disponible à l'époque.

C'est pourquoi le travail lexicographique du projet FEM a été réalisé en éditant des fichiers *Word* (sauvés en format RTF ou *Rich Text Format*), avec en général un fichier

par lettre initiale. Ces fichiers suivent un même modèle de document, dont chaque « style » (format d’affichage d’un paragraphe de texte *Word*) représente un élément dictionnaire (entrée, catégorie grammaticale, étiquette, glose, expression idiomatique, phrase exemple, prononciation, etc.). Ainsi, *Word* est utilisé comme un « pseudo-éditeur syntaxique ». (Voir *Figure 3 : une entrée du dictionnaire FEV*).

Durant le projet FEM, la forme de référence des données était une liste d’objets Common Lisp. Cette forme de référence était utilisée pour certains traitements, et on avait aussi développé une représentation dans une base de données relationnelle (*Ingres*), pour bénéficier des fonctionnalités associées (sauvegardes, retours en arrière, aspects transactionnels, vérifications diverses, statistiques...), très utiles pendant un développement intense et distribué. Cependant, depuis 2001, la forme de référence de FEM, et par suite du FEV, est constituée de fichiers XML, toutes les chaînes étant codées en Unicode (encodé en UTF-8).

La version informatique du FEM a été diffusée sur disquettes en 1996, avec des modes d’accès très puissants (application MCL (Macintosh Common Lisp) *Alex* de M. Lafourcade, données non piratables), puis en 1997-2001 sur Internet, via le serveur Silfide (action AUELF-UREF/CNRS) installé au GETA, puis en 1998 sur CD-ROM en HTML (données copiables mais non structurées), depuis 2001 sur le serveur du projet *Papillon* (<http://www.papillon-dictionary.org/>), et depuis avril 2003 sur CD-ROM multistandard sous forme d’une application *Java* (créée par G. Sérasset) donnant des modes d’accès presque aussi puissants qu’*Alex*, et totalement portable alors qu’*Alex* ne tournait que sur Macintosh.

Les données lexicales ont été revues à chaque nouvelle édition. Jusqu’à 2001, il manquait beaucoup d’équivalents anglais. M^{me} Lafourcade en a complété la plupart en 2001-2003. D’autre part, le contenu de la partie malaise a été considérablement revu par la MMM (Maison du monde malais, L. Metzger, La Rochelle) en 2002-2003, en accord avec le DBP, pour produire la dernière version. Pour le projet FEV, nous sommes partis de l’état du FEM en 2001, ce qui n’était pas gênant, puisque les seules améliorations apportées ensuite concernaient le malais et l’anglais.

3. PROBLÈMES DU TRAITEMENT INFORMATIQUE DU VIETNAMIEN

L’ancienne écriture vietnamienne utilise les caractères « Nôm » qui sont d’origine chinoise. Une transcription en caractères latins, dite Quốc Ngữ, a été proposée au XVII^e siècle par le missionnaire portugais Alexandre de Rhodes (1591-1660). Depuis lors, le vietnamien a connu de nombreux changements, et a été constamment développé. Avec plus de 80 millions de locuteurs, le vietnamien, langue tonale et monosyllabique, est actuellement une langue très importante en Asie du Sud-Est.

L’alphabet vietnamien s’écrit avec 17 consonnes simples (en minuscules *b c d đ g h k l m n p q r s t v x*) et 12 voyelles (*a â ã e ê i o ô o u y*). Le vietnamien dispose de six tons différents, chacun étant noté par un signe diacritique écrit toujours de la même façon, au-dessus ou au-dessous d’une voyelle. Par exemple, pour la voyelle *â*, on a : *â* (ton zéro ou absence de signe), *à* (ton remontant), *á* (ton descendant), *ã* (ton montant), *ã* (ton retombant) et *â* (ton intensif).

L'informatisation du vietnamien s'est d'abord concentrée sur les problèmes de base : codage des caractères, saisie sur un clavier ASCII, et restitution sur écran ou sur papier. Le fait qu'il faut au total coder 134 caractères avec leurs diacritiques a conduit à la proposition de nombreuses méthodes de codage. Il y a aujourd'hui au Viêt Nam une quarantaine de systèmes de codage (utilisant un ou deux octets) disponibles, dont certains sont très répandus (TCVN3-ABC, VNI, *Vietware*...). Cette situation pose le problème de la standardisation : à l'heure actuelle, on n'arrive toujours pas à se mettre d'accord pour utiliser un système de codage unique dans tout le pays, et il y a encore beaucoup de discussions sur l'utilisation d'Unicode.

Depuis 1996, l'ERTA (équipe de recherche en TA) de l'Université de Danang a construit et utilise une base de données lexicales sous forme d'une table de dimension 25 × 1087 dont chaque colonne représente une consonne simple ou composée et chaque ligne représente une syllabe tonique. Suivant le cas, chaque case de la table détermine un mot (formé par une consonne suivie d'une syllabe tonique) parmi plus de 7 000 mots usuels du vietnamien. Utilisant cette base, l'ERTA a développé des outils pour la correction orthographique, la construction des dictionnaires d'expressions ou d'orthographe, la conversion entre les systèmes de codage, etc.

4. DÉVELOPPEMENT D'UNE MÉTHODOLOGIE GÉNÉRIQUE

Dans le projet FEV, nous avons utilisé la forme XML définie pour le FEM comme forme d'échange. Pour le travail lexicographique « global », nous avons utilisé la forme « RTF stylé » déjà utilisée par le projet FEM. Nous avons aussi réalisé une partie du travail lexicographique sur des sous-ensembles de l'information (travail lexicographique « local ») : entrées simples, expressions et tournures idiomatiques, et exemples, en utilisant pour cela des fichiers *Word* (de structure différente) et des fichiers *Excel*. Par contre, nous n'avons pas utilisé d'objets (*Common Lisp* ou *Java*).

Pour faciliter le travail avec les fichiers *Word* et *Excel*, nous avons développé à Danang des programmes ou « macros » *Visual Basic (VBA)* automatisant la plupart des travaux répétitifs du projet : préparation des sources français-anglais, insertion des équivalents vietnamiens, conversion des codes et des formats RTF-XML, vérification automatique, etc. La multiplicité d'outils simples mais adaptés de cette méthode est une garantie de généralité, d'évolutivité et de réutilisabilité dans le champ de la production semi-automatique de dictionnaires électroniques multilingues.

Il y a plusieurs phases successives dans le processus de fabrication du dictionnaire FEV : prétraitement du FEM, choix d'un codage pivot, préparation des données à insérer, insertion automatique, insertion manuelle, et intégration des résultats.

4.1. Prétraitement du FEM

Les fichiers *Word* du FEM contiennent des équivalents malais, anglais, et (en partie) en thaï, dans une architecture « en éventail »¹. Le tableau ci-dessous donne le nombre de paragraphes à créer pour le vietnamien. Ils correspondent aux 3 styles *Word* associés aux entrées (*entry* ou *headword*), expressions ou groupes de mots (*phrase*) et phrases exemples (*sentence*).

Source FEM	Fichier	Taille (Koctets)	Nombre de paragraphes à traduire pour les styles :		
			Entry	Phrase	Sentence
FEM2A. src-1.72	A.doc	854	1 524	705	399
FEM2B. src-1.5	B.doc	616	938	452	155
FEM2C1. src-1.31	C1.doc	545	1220	712	145
FEM2C2. src-1.1	C2.doc	674	967	708	174
FEM2D. src-1.11	D.doc	842	1 405	645	304
FEM2E. src-1.11	E.doc	875	1 404	636	379
FEM2F. src-1.22	F.doc	479	836	592	239
FEM2G. src-1.31	G.doc	497	585	338	115
FEM2H. src-1.1	H.doc	269	432	249	59
FEM2I. src-1.22	I.doc	544	991	278	124
FEM2JK. src-1.32	JK.doc	157	217	158	30
FEM2L. src-1.22	L.doc	371	497	365	86
FEM2M. src-1.1	M.doc	643	1 168	792	134
FEM2N. src-1.2	N.doc	200	291	223	27
FEM2O. src-1.2	O.doc	205	379	254	43
FEM2P1. src-1.2	P1.doc	519	939	655	99
FEM2P2. src-1.1	P2.doc	538	926	726	103
FEM2Q. src-1.2	Q.doc	97	97	98	25
FEM2R. src-1.1	R.doc	730	1 291	676	123

¹ Au tout début de la construction du FEM, l'architecture était « en chaîne » : français-anglais-malais. Mais cela menait à des doublons quand on supprimait la vue de l'anglais. M. Lfourcade a alors opté pour l'architecture « en fourche » ou « en éventail » déjà utilisée en 1986 pour un dictionnaire français-anglais-japonais en télécommunications (contrat DGT-KDD-GETA-Champollion), et présentée à *Coling-86* à Bonn.

FEM2S. src-1.1	S.doc	625	1 298	803	100
FEM2T. src-1.1	T.doc	543	954	794	106
FEM2U. src-1.3	U.doc	68	83	52	7
FEM2V. src-1.21	V.doc	388	569	582	76
FEM2WXYZ. src-1.4	XYZ.doc	43	59	19	2
Total		11 322	19 070	11 512	3 054

Tableau 1 : répartition des éléments à compléter pour le vietnamien dans le FEM

Le modèle de documents FEM initial utilise des styles différents pour les éléments français, anglais et malais (fig. 2). Nous y avons donc ajouté de nouveaux styles pour le vietnamien.

Voici un résumé des tâches effectuées par les macros *VBA* dont nous avons parlé plus haut :

- Changer le nom de chaque fichier dans la base, par exemple, FEM2A. src-1.72 devient A.doc, etc. (fig.1.). Nous avons donc au total 24 fichiers de travail.
- Convertir tous les fichiers *Word* dans un même modèle de document contenant les nouveaux styles pour le vietnamien. Cacher tous les paragraphes des styles concernant le malais et le thaï.
- Supprimer les signes particuliers, comme le signe | dans *accepter*, etc. Les paragraphes concernés sont marqués pour pouvoir les restituer tels quels.
- Extraire du FeM trois parties distinctes et indépendantes : « Entry », « Phrase » et « Sentence », dont chacune contient 24 fichiers *Word* de même nom, de A.doc à XYZ.doc. Il y a donc 72 fichiers de travail.

Styles « français »	Styles « anglais »	Styles « malais »	Styles « vietnamiens »
Entry	english equ	malay equ	viet equ
french cat			
french entry number	english equ	malay equ	viet equ
french entry variant	english equ	malay equ	viet equ
french fem form			
french fem plur form			
french fem pron			
french gloss	english gloss	malay gloss	viet gloss
french label	english label		
french masc plur form			
french masc plur pron			
french phrase	english phrase	malay phrase	viet phrase
french plur form			

french plur form pron			
french plur form variant			
french plur form variant pron			
french pron			
french pron variant			
french sentence	english sentence	malay sentence	viet sentence
french subentry	english equ	malay equ	viet equ
french subentry pron			
cross ref entry			
cross ref marker			

Tableau 2 : styles du modèle de document *Word* utilisé pour le travail global sur le FEV

Voici une macro VBA permettant de convertir tous les fichiers dans un modèle unique :

```

Sub ToOwnModele()
' ToOwnModele Macro Enregistrée le 08/10/2001 par GRTA-DN
Dim i, k As Integer
Dim FeV(24), TxtABC(24)
FeV(1) = "A.doc"
...
FeV(24) = "WXYZ.doc"
For k = 1 To 24
  ChangeFileOpenDirectory "D:\FeV-2002\NewFeV0802\"
  Documents.Add Template:="FeV.dot", NewTemplate:=False, Document-
Type:=0
  ChangeFileOpenDirectory "D:\FeV-2000\FeMSource\"
  Selection.InsertFile FileName:=FeV(k)
  ChangeFileOpenDirectory "D:\FeV-2002\NewFeV0802\"
  ActiveDocument.SaveAs FileName:=FeV(k), FileFor-
mat:=wdFormatDocument, _
  LockComments:=False, Password:="", AddToRecentFiles:=True
  ActiveWindow.ActivePane.View.Zoom.Percentage = 100
  ActiveDocument.Save, ActiveDocument.Close
Next k
End Sub 'ToOwnModele

```

Figure 1 : macro de conversion dans un modèle unique

4.2 Choix d'un codage pivot

La deuxième phase détermine les systèmes de codage pour les caractères diacritiques vietnamiens. Il y a deux solutions qui coexistent : un codage de travail pour l'insertion et la vérification, et un codage pour le stockage et la conversion du dictionnaire.

Nous avons utilisé comme codage de travail TCVN3-ABC, qui est actuellement très répandu au Nord du Vietnam. Ce codage utilise les codes ASCII étendus (de 128 à 253 décimal) pour coder les caractères diacritiques comme « remplissages des vides ».

Pour le codage de stockage, nous avons utilisé le codage Telex qui n'utilise que des codes ASCII. Le principe du codage Telex est très simple : duplication ou accouplement des voyelles pour la représentation des signes diacritiques et utilisation des lettres *f, s, j, r, x* pour la représentation des tons. L'exemple ci-dessous présente l'utilisation de ces deux codages :

french_sentence	<i>cette mère a abandonné ses enfants</i>
TCVN3-ABC	<i>ngưòì mẹ này đã bỏ rơi những đũa con của mình</i>
Telex	<i>nguowowif mej nayf ddax bor rowi nhuwngx dduwas con cuar minhf</i>

L'utilisation des macros permet de convertir facilement à partir du codage Telex, qui joue le rôle de code pivot (figure 2), dans tous les systèmes de codages existant actuellement au Vietnam.

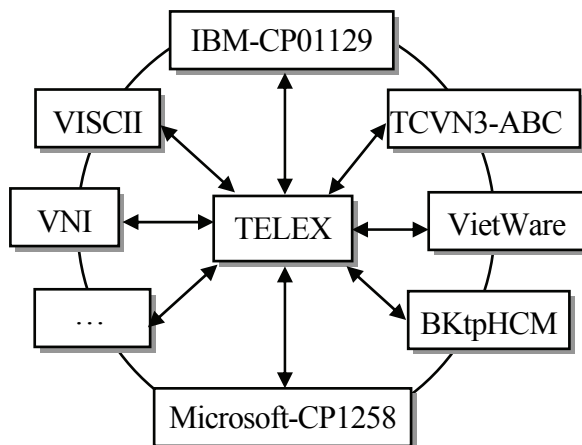


Figure 2 : code pivot Telex

4.3. Préparation des données à insérer

Dans cette phase, on construit une liste de paires. Le premier composant de chaque paire contient une rubrique en français à traduire en vietnamien, et joue le rôle « clé de travail ». Le second est son équivalent vietnamien, ou une liste d'acceptions correspondantes en vietnamien. Cette liste est contenue dans un ensemble de fichiers de documents *Word*. Chaque document est une suite de paires « clé-contenu » ayant la même nature de contenu. Nous avons donc des documents constitués, soit d'entrées (style « entry »), soit d'expressions (style « phrase »), soit de phrases exemples (style « sentence »). Pour enrichir la source d'insertion, nous prenons aussi les données en anglais.

Les données de tous ces documents sont complétées automatiquement ou manuellement par des équivalents vietnamiens à partir de plusieurs sources dictionnaires bilingues (français, vietnamien et anglais) disponibles actuellement sous forme de dictionnaires informatisés, comme le *LacViet*, ou en ligne sur le Web. La récupération de ces équivalents est effectuée soit par des microcaptures d'écran, soit par des macros de transformation écrites en *VBA*. Il y a deux étapes à réaliser : accumulation des données brutes et hétérogènes provenant de différentes ressources, puis raffinement selon les acceptions déterminées par les gloses figurant dans les entrées du dictionnaire FEM.

4.4 Insertion automatique

Dans cette phase, il y a trois sous-tâches correspondant à l'insertion des équivalents vietnamiens dans les trois groupes de fichiers *Word*, « Entry », « Phrase » et « Sentence » (24 pour chacun). L'insertion est effectuée d'une manière distincte et indépendante. Le fonctionnement des macros d'insertion automatique est le suivant :

- Ouvrir la source d'insertion suivant le style de travail : « entry », « french_phrase », ou « french_sentence ».
- Ouvrir le fichier FEM global correspondant.
- Parcourir de haut en bas ce document et traiter successivement les données associées à chaque « clé de travail », *i.e.* à chaque acception française. Ces données sont constituées d'une suite de paragraphes *Word*, avec un paragraphe « stylé » pour chaque information élémentaire.
- Créer automatiquement dans le document FEM un nouveau paragraphe « vietnamien » du style correspondant au type d'information à insérer.
- Chercher la « clé de travail » dans la source d'insertion courante. Si on la trouve, on insère dans le paragraphe qu'on vient de créer l'équivalent vietnamien trouvé. S'il y en a plusieurs, on crée un nouveau paragraphe pour chacun. Si la clé n'est pas trouvée, on insère dans le paragraphe créé une marque, par exemple « <<<< », pour indiquer qu'il y a un « trou » dans l'information sur le vietnamien.
- À la fin, enregistrer ce document dans un autre fichier *Word* comme résultat FEV.
- Faire une analyse statistique en même temps et marquer les paragraphes-clés qui n'ont pas d'équivalents vietnamiens.

L'exécution des macros concernées est effectuée alternativement et séparément en plusieurs fois. On donne priorité aux données en français, l'insertion sur les données en anglais est donc exécutée à la fin. Les fichiers partiels FEV correspondant à un même ensemble d'entrées sont consolidés en un fichier unique afin d'obtenir un fichier complet avant l'étape d'insertion manuelle. La répartition est la suivante :

Tâches d'insertion	Source FEM	Pourcentage
Entry	19 070	67 %
Phrase	11 512	54 %
Sentence	3 054	100 %

Table 1 : Répartition des tâches d'insertion

4.5. Insertion manuelle

L'insertion automatique ne suffit pas pour achever complètement la construction du dictionnaire. Il y a encore des vides à combler qui demandent une insertion manuelle. Ce travail a été réalisé par certains de nos étudiants connaissant le français et/ou l'anglais (rarement les deux), et par des enseignants de français bénévoles de l'Université de Danang. Utilisant en même temps la source d'insertion et les dictionnaires bilingues (français, anglais, vietnamien) disponibles, le travail de saisie a été fait sur des IBM-PC, toujours sous *Word*™. À raison de 7 à 8 entrées par heure, environ 3 500 à 4 500 heures de travail ont été effectuées.

4.6. Intégration des résultats

Cette phase réalise l'intégration des résultats produits dans les phases précédentes. Elle consiste à extraire les équivalents vietnamiens des documents *Word* provenant de la division en trois parties distinctes et indépendantes et à les insérer dans les documents « maîtres ».

Partant des fichiers du style « Entry », qui sont ouverts comme documents principaux, la macro d'intégration prend chaque nouveau paragraphe dans le document partiel approprié et l'insère à la position correcte. La clé de recherche permet de déterminer cette position. Le document dans lequel est effectuée l'intégration des équivalents vietnamiens est parcouru deux fois de haut en bas, d'abord pour le style « viet_phrase », puis pour le style « viet_sentence ».

Une fois cette phase d'intégration effectuée, on dispose d'une version « opérationnelle » du dictionnaire FEV, sous forme d'un ensemble de documents *Word*. À titre d'exemple, la Figure 3 ci-dessous en donne une entrée.

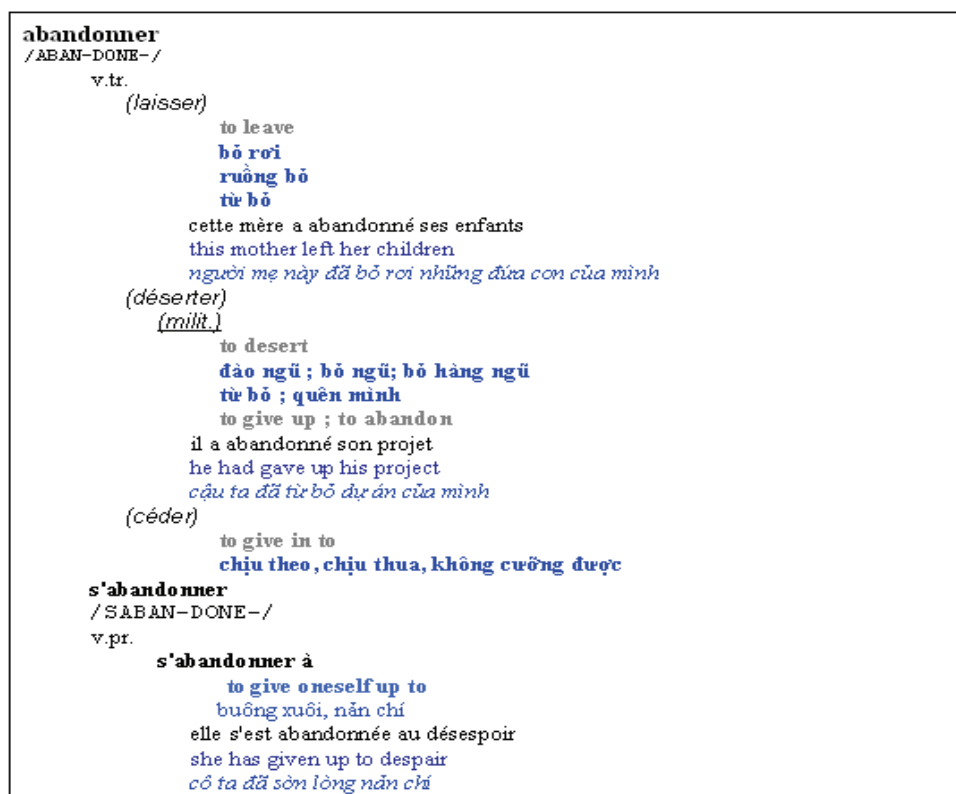


Figure 3 : une entrée du dictionnaire FEV

4.7. Vérification et correction

La vérification et la correction linguistique sur les documents *Word* ont été effectuées manuellement à Danang par nos étudiants et enseignants de français. Cette phase a demandé un temps de travail considérable (environ 60 entrées à l'heure, soit environ 400 heures de travail) et a nécessité la collaboration d'une personne très compétente en français et en vietnamien.

Le schéma du processus de construction du FEV peut être résumé par la figure suivante.

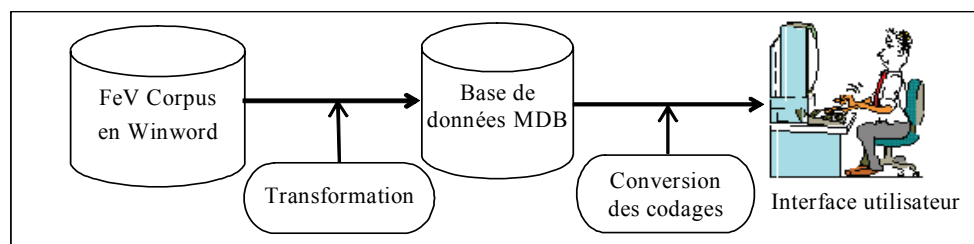


Figure 4 : étapes du processus de création du dictionnaire FEV

5. UTILISATIONS DU FEV

5.1. Mise du FEV sur la plate-forme *Papillon* en mode consultation

La partie « CDM » (*common dictionary markup*) de la plate-forme *Papillon* permet non seulement de consulter tout type de dictionnaire au format XML, grâce à une interface générique de consultation, mais aussi d'éditer et de réviser en ligne les contributions sur des articles de dictionnaire, quelle que soit leur structure, à condition de développer (à partir d'une forme générique) l'interface correspondante.

Nous avons donc importé le FEV sur cette plate-forme (<http://www.papillon-dictionary.org/>), ce qui permet de le consulter en ligne, seul ou en parallèle avec les autres dictionnaires disponibles. Cet import n'a pas posé de problème particulier, puisque le dictionnaire FEM avait une structure similaire. Il a suffi de transformer le FEV en XML avec les mêmes programmes que ceux utilisés pour le FEM.

Nous prévoyons d'utiliser cette plate-forme pour faire ensuite une révision des équivalents vietnamiens, à la *wiki*, à travers le web. La plate-forme *Papillon* gérant des utilisateurs et des droits d'accès, nous pourrions éventuellement organiser une partie du travail en constituant des groupes de réviseurs, par exemple des groupes d'étudiants vietnamiens devant participer à cette révision dans le cadre de leur étude du français. Nous savons que cela est possible et fonctionne très bien, puisque le site IToldU a permis de collecter plus de 17 000 termes anglais-français, avec environ 12 000 exemples d'usage corrects, dans un contexte similaire. Il faudra pour cela intégrer à *Papillon-CDM* une interface simple et spécifique de la structure en fourche utilisée par le FEM et le FEV.

5.2. Applications spécifiques construites autour du dictionnaire FEV

À partir des fichiers du dictionnaire FEV, nous avons créé des dictionnaires de consultation sous plusieurs formes :

- par fenêtrage,
- par « cliquer-et-voir »,
- sous forme d'interface Web,
- un CD-ROM en HTML,
- et une version papier.

5.3. Autres actions prévues autour du FEV

Nos perspectives sont de continuer le travail coopératif dans le cadre du projet international *Papillon* (création collaborative de bases lexicales multilingues avec formalisme pivot), pour développer de nouveaux dictionnaires : VEF, et VEL* reliant le vietnamien à d'autres langues de la Francophonie. Pour cela, il faudra « mettre le vietnamien » dans la partie *Papillon-Nadia* (dictionnaires monolingues au format DiCo de Polguère-Mel'čuk).

On commencera par définir la microstructure DiCo du vietnamien au moyen d'un schéma XML dérivé des schémas déjà disponibles pour les autres langues (français, anglais, japonais, malais...). Ensuite, on importera les données du FEV de *Papillon-CDM* dans *Papillon-Nadia*. Cela nous donnera des entrées « squelettes », comportant une seule lexie par vocable vietnamien. En traitant le graphe des liens traductionnels bilingues par différents algorithmes, on obtiendra un « premier jet » du « graphe Nadia », c'est-à-dire une division des vocables en lexies, et des liens interlingues (« axes ») reliant les lexies de différentes langues, dont le vietnamien.

Une fois qu'on aura ce premier jet, c'est-à-dire les volumes DiCo pour le vietnamien, le français, l'anglais, etc. ainsi que les axes (liens interlingues) reliant les lexies « traditionnellement équivalentes », on pourra extraire de la base *Papillon* des dictionnaires partant du vietnamien et allant vers le vietnamien, vers et depuis toutes les autres langues disponibles.

Au début, seuls les dictionnaires partant de langues pour lesquelles on avait déjà des divisions en sens, et donc des lexies, seront de qualité équivalente à celle des dictionnaires initiaux. Ensuite, le travail des contributeurs devrait améliorer la division des vocables en lexies, et le graphe des lexies et axes. De plus, des informations plus détaillées devraient être ajoutées dans le DiCo vietnamien comme dans les autres DiCos : formule sémantique, régime, valeurs de nombreuses FLS (fonctions lexicosémantiques), exemples d'usage, collocations libres.

6. CONCLUSIONS ET PERSPECTIVES

Le projet FEV a permis d'obtenir un dictionnaire français-vietnamien informatisé d'environ 20 000 entrées, 50 000 acceptions, avec quelque 10 000 exemples et 8 000 tournures. Il est disponible sur le web en consultation, et plusieurs outils dérivés ont été réalisés. Cela a été possible grâce à la réalisation préalable, d'une part d'un dictionnaire comparable FEM (déjà complètement finalisé, en trois versions successives depuis 1996), sur conventions MAE, et d'autre part d'une partie d'un FET (français-thaï via l'anglais, dont 12 % ont été réalisés et dont le financement s'est arrêté à cause de la crise en Thaïlande en 1998-2000).

La qualité actuelle correspond à un premier niveau, et il faudra encore beaucoup de travail pour obtenir une qualité similaire à celle du FEM. L'étape de révision suivante devrait pouvoir être effectuée par des contributeurs travaillant directement sur la base de données sous-jacente, sur la Toile, en utilisant les navigateurs disponibles.

Le projet FEV a encore une fois démontré la possibilité de faire le travail « de base » avec des étudiants de faible niveau de français et/ou anglais, puis de le compléter pour les parties plus difficiles avec des spécialistes, en l'occurrence des professeurs de français de l'Université de Danang. Au plan pratique, notre solution permet la réalisation de dictionnaires papier et informatisés qui font actuellement défaut, et ce dans différents formats, avec possibilité de filtrer certaines informations.

Nous avons effectué ce travail à l'aide d'une multiplicité d'outils simples mais adaptés, réalisés sous forme de macros *VBA* appelées depuis *Word*, utilisé pour le travail lexicographique comme un pseudo-éditeur syntaxique.

Dans le futur, notre projet est de « symétriser » les données en divisant les vocables vietnamiens en lexies (sens de mots), et ainsi d'insérer le vietnamien dans la base lexicale multilingue par acceptions interlingues *Papillon*, ouvrant la voie à la production de dictionnaires bilingues ou multiculturels entre le vietnamien et les autres langues de la base *Papillon* (il y en a déjà 9, et ce nombre est destiné à croître).

Nous espérons que cette approche pourra être reprise pour d'autres langues de la francophonie et contribuera à la construction d'outils (en particulier d'outils de traduction automatique ou automatisée) reliant le vietnamien au français, et plus généralement aux langues « informatisées » et « équipées » présentes sur la Toile.

BIBLIOGRAPHIE

- Doan (N. H.), 1998 : « Accumulation of Lexical Sets : Acquisition of Dictionary Resources and Production of New Lexical Sets », 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics, *Proc. Coling-ACL '98*, vol. I, Montréal, Canada, 10-14/8/1998, pp. 330-335.
- Fafiotte (G.) et Boitet (C.), 2002 : *Réalisation d'un dictionnaire d'usage et d'une base terminologique par acceptions informatisées français-vietnamien via l'anglais*, rapport final du projet FEV, GETA, CLIPS, IMAG (UJF, CNRS et INPG), Grenoble, 19 pp.
- Lafourcade (M.), 1996 : « Serveurs de dictionnaires – Étude de cas avec l'outil Alex et le projet de dictionnaire français-anglais-malais », *actes séminaire Lexique*, Grenoble, 13-14/11/1996, CLIPS-IMAG, pôles langage naturel et parole du GDR-PRC CHM., vol. I, pp. 185-192.
- Lafourcade (M.), 1997 : *Construction et services de dictionnaires n-lingues, exemple des projets Fe**, quatrième conférence annuelle sur le traitement automatique du langage naturel, 12-13/06/1997, Grenoble, vol. I, pp. 162-168.
- Phan (H. K.) et Boitet (C.), 1992 : « Multilingualization of an editor for structured documents. Application to a trilingual dictionary » *Proceedings of the 15th International Conference on Computational Linguistics*, Coling-92, vol. III/IV, Nantes, 23-28/08/1992, pp. 966-970.
- Phan (H. K.), 2002 : « A Survey on Vietnamese Language Processing and Multilingual Processing », *Proceedings of Nectec Conference*, Pukhet, Thailand, 3-5/10/2002, pp. 333-337.
- Vo-Trung (H.), 2004 : *Méthodes et outils pour utilisateurs, développeurs et traducteurs de logiciels en contexte multilingue*, thèse en informatique, INPG.

Sites web :

<http://www.informatik.uni-leipzig.de/~duc/software/Dict/vietdict.html>

<http://www.vietlex.com.vn/index.htm>

<http://www.chez.com/halongbay/vietphap.htm>

<http://www.freelang.com/dictionnaire/vietnamien.html>

<http://dict.vietfun.com/>

<http://www.lacviet.com.vn>

<http://www.papillon-dictionary.org>

<http://opus.grenet.fr/itoldu/ITOLDU>

LES CONTEXTES DE L'ENQUÊTE : LE CAS DE L'ATLAS LINGUISTIQUE DE TUNISIE

Soundous Ben Hariz Ouenniche
Université de la Manouba
Manouba, Tunisie

Résumé : *Le propre d'une enquête linguistique est d'amener l'informateur à énoncer un mot cible qui ne doit en aucun cas être fourni par un tiers. L'enquêteur se doit donc de préparer un contexte linguistique afin de mettre en situation l'informateur. L'élaboration des contextes de l'enquête doit tenir compte de paramètres variables : le type de questionnaire, la relation signe-référent, la typicité ou la saillance des propriétés et l'univers de croyance.*

Mots-clés : *mot, contexte linguistique, enquête, informateur, questionnaire, enquêteur*

1. INTRODUCTION

Les enquêtes de l'*Atlas linguistique de Tunisie* (ALT) ont été effectuées en trois étapes à partir de trois types de questionnaires différents : phonologique, morphosyntaxique et lexical. Dans l'ensemble, ces enquêtes ont touché 3 000 informateurs, soit quatre types d'informateurs par point d'enquête (homme/femme adultes ; homme/femme âgés). Ces points d'enquête correspondant à l'organisation administrative et territoriale du pays sont au nombre de 250.

Pour garantir l'objectivité des résultats, nos enquêteurs avaient comme consigne d'amener l'informateur à énoncer un mot cible qui ne doit en aucun cas être fourni par un tiers. Ils se devaient donc de concevoir des contextes précis qui permettent à l'informateur d'énoncer le mot recherché. Mener à bien une telle tâche est fonction de paramètres variables, puisque :

1. Le contexte dépend tout d'abord du type de questionnaire : dans le cadre d'un questionnaire phonétique, l'information recherchée concernera le signifiant ou la matière vocalique. S'il s'agit d'un questionnaire de morphosyntaxe, elle traitera de l'articulation au niveau syntactico-sémantique. Cela peut concerner simplement un fait sémantique pour un questionnaire lexical.
2. Le contexte dépend aussi du degré de complexité dans la relation signe-référent : ainsi, une même dénomination peut désigner des référents différents. De même qu'un objet précis peut avoir plusieurs dénominations d'une région à une autre.
3. Il dépend, d'autre part, de la typicité ou de la saillance des propriétés qu'il évoque : certains référents ne peuvent pas être atteints à partir de leurs propriétés analytiques, l'élaborateur du contexte se trouve alors dans l'obligation d'opérer des choix à

l'intérieur d'une gamme de propriétés sujettes à des variations (usage, forme, fonction, relation à l'homme...).

4. Enfin, l'élaboration des contextes dépend aussi de l'univers de croyance qui peut être différent entre un enquêteur et un informateur ou même d'un informateur à un autre.

Ainsi, le contexte, en tant qu'outil d'analyse linguistique, joue-t-il un rôle déterminant dans la structuration des enquêtes. C'est cette valeur heuristique que nous comptons expliciter en nous appuyant sur des exemples précis tirés de l'expérience de *L'Atlas linguistique de Tunisie*, qui a pour objectif, rappelons-le, de décrire le système dialectal tunisien.

2. LES CONTEXTES ET LES TYPES DE QUESTIONNAIRES

La conception d'un contexte varie selon la nature du fait linguistique visé :

2.1. Le questionnaire phonétique

La dimension phonologique a pour objectif, dans le cadre d'une enquête linguistique, de « dégager les contours du système phonologique [...], son fonctionnement et son économie. »¹ Les faits phonologiques sont appréhendés à travers des unités lexicales. L'objectif de l'enquêteur est donc de faire dire une unité lexicale précise contenant le phonème voulu.

2.1.1. L'opposition [g]/[q]

À titre d'exemple prenons, dans le système consonantique de l'arabe dialectal tunisien, la fameuse opposition [g]/[q], qui a toujours été considérée comme la transposition de traits phonétiques opposant les parlers bédouins aux parlers citadins. Les variations régionales détectées ont montré qu'il s'agit d'un phénomène plus complexe. Dans les régions où l'on prononce pertinemment le phonème [q], un même mot peut être prononcé différemment par un même locuteur mais dans deux contextes distincts : dans le premier contexte, il s'agit d'un emploi libre, dans le second d'un emploi figé.

Par exemple, pour l'adjectif de couleur *bleu* au féminin singulier زرقة [zarqa], l'enquêteur est obligé de concevoir deux contextes : dans le premier, l'enquêteur incitera l'informateur à donner la couleur d'un objet quelconque, ici en l'occurrence la couleur bleue. Dans le second, il devra vérifier la prononciation de cet adjectif dans un syntagme figé où la base est [ʕi:n] عين, *œil*. En effet, là où en emploi libre on dit « ورقة » [warqa zarqa] *une feuille bleue*, « كرهية زرقة » [karahba zarqa] *une voiture bleue*, « كرة زرقة » [kura zarqa] *un ballon bleu*, les informateurs prononcent systématiquement عين الزرقة [ʕi:n zarga] *un œil bleu*.

¹ Baccouche T. et Mejri S. 1998, *L'Atlas linguistique de Tunisie, problématique phonologique*, AIDA III, Malta.

Le même fait est constaté pour le mot قلب [qlab], diminutif de cœur قلب [qalb]. Quand le diminutif réfère à un objet du monde, il est prononcé قلب [qlayeb], quand il est utilisé dans un syntagme prépositionnel à signification métaphorique : *avec conviction*, il est prononcé قلب [glayeb]: يخدم بالقلب [yaɣdem bilglayeb], *il travaille* « *avec conviction* ».

2.2. Le questionnaire morphosyntaxique

Pour une description du système linguistique dialectal tunisien, l'enquêteur est amené à faire des choix sur le fait linguistique à étudier. Dans le cadre du questionnaire morphosyntaxique, la cible est la structure de la phrase et l'information recherchée concernera l'articulation au niveau syntactico-sémantique.

L'enquêteur est amené à présenter une série de questions qui permettront de rendre compte de la morphologie et de la syntaxe des unités formant la phrase dialectal tunisien. Le contexte est ainsi appréhendé comme une investigation linguistique.

Certaines informations recherchées impliquent l'énonciation de paradigmes élaborés qui nécessiteraient un certain niveau d'instruction. Par exemple, il est difficile d'imaginer un analphabète énoncer les paradigmes de flexions verbales des 23 verbes contenus dans le questionnaire morphosyntaxique et ce à l'accompli, l'inaccompli et à l'impératif.

Pour rendre compte d'une telle difficulté, nous présentons le contexte suivant, élaboré par l'enquêteur lors de la pré-enquête. (L'informateur étant un homme âgé de plus de 60 ans) :

توا يا سيدي باش نعملو تصريف

Littéralement : *Maintenant monsieur, nous allons faire de la conjugaison*

اه

Littéralement : *Ah !*

في الماضي والمضارع و الأمر

Littéralement : *À l'accompli, l'inaccompli et l'impératif*

لا مانفهمش حتى شيء

Littéralement : *Non, je ne comprends rien.*

La première réaction de l'informateur à l'annonce d'une série de questions portant sur la conjugaison témoigne de l'importance du contexte linguistique : l'utilisation de jargon métalinguistique face à cet informateur illettré n'a fait que le rebuter et peut même bloquer la suite du questionnaire. Face à cette situation, l'enquêteur change d'angle d'attaque en envisageant un scénario où l'informateur se trouve impliqué dans un rôle bien déterminé. Il demande à celui-ci de jouer le rôle d'un père qui a frappé son fils et à partir de là, il génère la suite du paradigme en changeant à chaque fois la personne ou le temps :

لا وراسك تفهم ايا تقول انا ضربت ولدي انت

Littéralement : *Non, vous comprenez je le jure sur votre tête, allons-y, vous dites j'ai frappé mon fils, tu...*

ضربت

Littéralement : *frappes*

هو

Littéralement : *lui (il)*

ضربو

Littéralement : *frappe*

Et ainsi de suite...

2.3. Le questionnaire lexical

Si la phonétique et la morphosyntaxe sont conditionnées par le caractère fermé de leur système, il n'en est pas de même pour le lexique qui se définit par son caractère ouvert, ce qui ne facilite pas la tâche de l'enquêteur. En effet, celui-ci est amené à varier les contextes afin de rendre compte de toute la richesse du lexique.

Se contenter d'une seule requête risque de donner lieu à une saisie parcellaire de la variété de la dénomination lexicale.

Soit le contexte suivant :

أنا عندي برشه مشاكل. بت ليلة كاملة و أنا ؟

Littéralement : *J'ai beaucoup de problèmes, j'ai passé la nuit entière...*

تفكري

Littéralement : *Vous réfléchissez*

معناها؟

Littéralement : *Ce qui veut signifie ?*

راك دايدة و تخممي

Littéralement : *Vous êtes sonné, vous réfléchissez*

والا

Littéralement : *Ou bien ?*

فلان مش يضرب خماسه في سداسه

Littéralement : Il multiplie son un cinquième par son un sixième.

À partir d'un scénario où sont définis un objet (avoir des problèmes), un rôle (présenté par le pronom personnel « je ») et un circonstant (passer toute une nuit), l'enquêteur réussit à recueillir une première réponse تفكري [tfakri], *vous réfléchissez*, puis à partir des tournures métalinguistiques معناها [maʕnaha], *ce qui signifie*, et والا [wa illa], *ou bien*, il amène l'informateur à donner d'autres synonymes libres : دايدة [dayɣa] *vous êtes sonné*, تخممي [tɣamammi] *vous réfléchissez*, ou figé: فلان مش يضرب خماسه في سداسه: [flɛn mʃa yadɾab ɣmassu fi sdassu] *Il multiplie son un cinquième par son un sixième*.

Dans d'autres cas, la dénomination peut être source de confusion ou de variation pour des raisons socio-environnementales. La précision de la réponse recherchée dépend de l'univocité du contexte : l'exemple type est celui des noms de poisson qui ne relèvent pas de la compétence de l'ensemble des locuteurs (les habitants de la côte et ceux de l'intérieur du pays). D'où l'efficacité des supports iconographiques.

3. CONTEXTE ET VARIATION DÉNOMINATIVE

Le degré de complexité dans la relation signe-référent est un paramètre important qui doit être pris en considération lors de la préparation des contextes des questionnaires linguistiques : si une même dénomination peut désigner des référents différents, le cas inverse existe également et un objet précis peut avoir ainsi plusieurs dénominations d'une région à une autre.

3.1. Un référent, plusieurs dénominations

Les abricots, par exemple, ont deux dénominations, مشماش [maʃmaʃ] et برفوق [bargug] : Cette dernière dénomination, [bargug] برفوق, a une valeur générique à Gafsa, autre-

ment dit, elle dénomme l'espèce. Ailleurs, c'est مشماش [maʃmaʃ] qui est générique et برفوق [bargug] est une sous-catégorie, désignant une variété d'abricot de petite taille².

Le mot برفوق [bargug] étant présent sur tout le territoire, l'enquêteur doit décrire avec précision le référent qu'il vise afin de s'assurer de la signification relative à chaque dénomination.

3.2. Une dénomination, des référents différents

L'élaboration des contextes de l'enquête doit donc prendre en considération les variantes dénominatives, mais elle se doit aussi de prendre garde à la différence référentielle : en effet, une dénomination peut avoir différents référents d'une région à une autre, ce qui témoigne de l'hétérogénéité du système référentiel.

Le mot شمام [ʃɛ:m] se réfère selon les régions à trois objets différents :

- Une espèce de végétation à odeur parfumée dans la région du Sahel ;
- Une sorte de melon dans la région de Kébili ;
- Des oranges à Gafsa.

Si l'enquêteur cible un seul de ces trois référents, il risque de ne pas obtenir le signifiant voulu. Donc pour le mettre en évidence, il doit envisager trois contextes relatifs aux trois référents.

4. CONTEXTE ET VARIATION DES REPRÉSENTATIONS

Le troisième paramètre concerne celui de la typicité ou de la saillance des propriétés : il s'agit d'objets qui ne peuvent pas être appréhendés à partir de leurs seules propriétés analytiques. Dans ce cas, leur reconnaissance n'est possible qu'à travers des représentations plus ou moins variables.

Soit par exemple, l'ustensile باقية [ba:gja]. Sa définition minimale peut être quelque chose comme un petit ustensile en terre cuite. Selon les régions, il est distingué soit par sa forme plus ou moins allongée, soit par l'usage : il s'agit d'un récipient destiné à contenir un quelconque aliment, ou encore le *henné* ou le *harqous*, sorte de liquide noir avec lequel les femmes se font des tatouages, ou décorent le henné enduit sur les mains.

5. CONTEXTE ET UNIVERS DE CROYANCE

L'élaboration des contextes linguistiques doit prendre en considération l'univers de croyance qui peut être différent entre un enquêteur et un informateur ou même d'un informateur à un autre.

² Dans certaines régions, برفوق [bargug] se réfère à une variété de prune.

Il arrive que lors des enquêtes certains mots mis en contexte choquent les informateurs. Citons à ce propos [goʃr] « goujat », [manqu:ba] « trouée », [mʃalgha] « étendu ou assis avec nonchalance ». Dans certaines régions, ces mots sont connotés négativement ou même sexuellement.

Il en est de même pour certaines expressions figées dont la charge sémantique péjorative varie d'un point d'enquête à l'autre. À ce propos, nous aimerions citer l'exemple d'une situation inconfortable dans laquelle s'est retrouvé l'un de nos enquêteurs en voulant inciter l'informateur à prononcer l'adverbial « six douze », qui, précédé du syntagme verbal [drabha] (littéralement : *frappé*) semble signifier exclusivement « se saouler », ce qui rime avec interdit :

« Par exemple, on dit ta femme *darbatha six douze bel kosksi* » (littéralement : *On dit de ta femme qu'elle est frappée six douze avec le coucou*) L'informateur, énervé, lui répond littéralement : « on n'a pas de femme qui se comporte de cette façon indigne ».

Pour l'informateur, l'intensif « six douze » est figé, il ne s'applique qu'à un contexte bien déterminé.³

La notion d'univers de croyance est à prendre en compte lorsque les enquêteurs sont amenés à parcourir différentes régions qui témoignent d'un certain conservatisme moral.

Le mot [tmanjak] qui signifie « se moquer de quelqu'un » et plus familièrement « se foutre de la gueule de quelqu'un », est très marqué dans certaines régions du nord ouest de Tunis ou du sud puisqu'il a une connotation sexuelle, comme peut le montrer la racine du terme.

6. CONCLUSION

L'*Atlas linguistique* est une investigation linguistique de terrain ayant pour objectif de collecter des données relatives au parler tunisien afin de décrire le système qui lui est propre.

Afin de mener à bien cette investigation linguistique, le contexte s'avère être un outil primordial. Mais il n'y a pas de recette toute faite, le contexte est le produit de plusieurs paramètres. Il est le résultat aussi d'une interaction entre l'information recherchée et l'information obtenue, la situation envisagée et la situation réelle de l'enquête...

Prendre en considération tous ces paramètres permettrait la réalisation d'un atlas illustrant le système linguistique du dialectal tunisien. Cette réalisation serait d'autant plus intéressante si elle était accompagnée d'un atlas parlant, c'est ce à quoi toute l'équipe de l'*Atlas linguistique de Tunisie* travaille en collaboration avec le LLI.

³ L. Oueslati, « L'enquête entre validation des questionnaires et qualité des données collectées », *Actes du Colloque International AIDA 6*, Tunis (à paraître)

BIBLIOGRAPHIE

- Baccouche (T.) et Mejri (S.), 1998 : « L'Atlas linguistique de Tunisie: problématique phonologique », Aida III, Malta.
- Baccouche (T.) et Mejri (S.), 2000 : « L'Atlas linguistique de Tunisie: spécificités phonologiques », Aida IV, Marrakech.
- Boisson (Cl.) et Thoiron (Ph.), 1997 : *Autour de la dénomination*, Lyon, PUL.
- Gross (G.), 1996 : *Les Expressions figées en français*, Paris, Ophrys.
- Mejri (S.), 1997 : *Le Figement lexical*, faculté des lettres de La Manouba, Tunis.
- Mejri (S.), 2000 : « L'Atlas linguistique de Tunisie : le questionnaire morphosyntaxique », Aida IV, Marrakech.
- Oueslati (L.), 1999 : « De l'identique au différent, le parcours d'un euphémisme », VI^{es} journées scientifiques du réseau LTT, *L'Éloge de la différence : la voix de l'Autre*, Beyrouth.
- Oueslati (L.) : *L'enquête entre validation des questionnaires et qualité des données collectées*, Aida VI 2004, Tunis (à paraître).
- Ouerhani (B.), 1999 : « Polylexicalité ou quand la voix de l'autre se fige dans les mots », VI^{es} journées scientifiques du réseau LTT, *L'Éloge de la différence : la voix de l'Autre*, Beyrouth.
- Said (M.), 1999 : « L'emprunt, la difficile intégration de l'autre » dans VI^{es} journées scientifiques du réseau LTT, *L'Éloge de la différence : la voix de l'Autre*, Beyrouth.
- Sandly (S.), 2000 : « L'Atlas linguistique de Tunisie : pratiques de l'enquête », Aida IV, Marrakech.

MOTS ET CONTEXTES EN FPI ET EN NOUCHI

Noël Kouassi Ayewa

département des sciences du langage

UFR-LLC/Université de Cocody

Abidjan, Côte d'Ivoire.

Résumé : *Les variétés de français de Côte d'Ivoire ont forgé leur lexique à partir de la déformation des mots français, de l'alliage des mots des langues européennes avec des mots des langues ivoiriennes mais aussi de l'emprunt de mots du terroir ou des mots du français enrichis de sens nouveaux.*

Des dictionnaires traitent de leur lexique : mais comment, à l'aide de la métalangue, présenter des réalités culturelles inconnues à des étrangers ? L'image et le son ne seraient-ils pas nécessaires ?

Mots-clés : *français populaire ivoirien, nouchi, dictionnaire, son, image*

La langue est par excellence un moyen de communication entre les membres d'une communauté donnée. La création lexicale est l'une des fonctions essentielles de l'activité langagière grâce à laquelle une langue assure la totalité de l'expérience humaine. Elle permet, en cas de nécessité, d'étendre l'expérience linguistique aux limites nouvelles de l'expérience humaine en inventant des mots nouveaux. Ainsi, avec l'invention des hydravions au début du XX^e siècle, le mot *amerrir* apparaît en français en 1912¹, plus tard, en 1921, la pensée de l'être humain s'engageait dans un voyage interplanétaire et alunissait en parachutage sur la lune, d'où le mot *alunir*². Ces deux mots sont mis en vogue à partir du 1^{er} septembre 1969 quand l'astronaute américain Neil Armstrong foula pour la première fois de l'histoire de l'humanité le sol lunaire, et entreprit de regagner la terre en plongeant son engin dans l'océan Atlantique.

Ces deux mots sont des parasyntèses créées sur le modèle d'atterrir. Ils comportent chacun un préfixe *a-*, et un suffixe *-ir* qui classe ces deux verbes dans le deuxième groupe. Ces affixes s'adjoignent à une base lexicale qui indique quant à elle, sur le modèle d'*atterrir*, le lieu qui va accueillir un avion qui atterrit quand il quitte l'espace aérien pour se poser sur la terre ; et *alunir*, c'est poser une navette spatiale sur la lune comme *amerrir*, c'est également faire poser un engin spatial sur la surface de la mer. Ces mots créés selon le génie de la langue française n'étonnent personne ; ils s'intègrent donc parfaitement à la langue française. Par contre, il existe des créations lexicales qui peuvent choquer car elles n'obéissent pas aux principes de création lexi-

¹ Robert : *Le petit Robert*, cf. *amerrir*.

² Robert : *op. cit.*, cf. *alunir*.

cale d'une seule langue : c'est le cas de la création lexicale en milieu diglossique, plus précisément en « français populaire ivoirien » (FPI) ou en « nouchi ».

Comment donc sont créés les mots dans ces variétés de français ? Ces mots, bien que présentant souvent une consonance française, sont-ils accessibles à tout francophone ? En effet, il a été consacré des dictionnaires du type classique aux mots de ces variétés ivoiriennes du français ; mais peut-on arriver au moyen de la métalangue uniquement à véhiculer les sens de ces mots inspirés par un contexte socioculturel spécifique ? Autrement dit, n'est-il pas nécessaire aujourd'hui d'étayer ces définitions d'images et de sons qui rendraient compte fidèlement de la réalité socioculturelle qu'expriment ces mots ? Telle est la problématique que nous abordons dans le présent exposé.

1. LA CRÉATION LEXICALE EN FPI

Les mots du FPI proviennent en majorité de la déformation phonologique et morphologique des mots français ; toutefois, le FPI use de procédés spécifiques tels que l'emprunt, la troncation... (cf. Annexe I).

1.1. Déformation des mots français

Tous les mots français renfermant les voyelles orales centrales [ə, Ø] ainsi que les consonnes palato-alvéolaires, surtout la sourde [ʃ], sont systématiquement escamotés en FPI ; ainsi obtient-on par déformation des mots français les mots suivants :

- [ʃ] * *tousse* au lieu de « touche » (cf. [4])
 - * *seque*, « chèque » (cf. [5])
 - * *sarse*, « cherche » (cf. [12])
 - * *sampion*, « champion » (cf. [12])
- [Z] * *auzoud'hui* au lieu de « aujourd'hui » (cf. [5])
 - * *couraze*, « courage » (cf. [8 et 9])
- [ə] * *pétit* au lieu de « petit » (cf. [3.b])
 - * *véni*, « venu » (cf. [3.b])
 - * *yé*, « je » (cf. [4.a et 5.b])
- [Ø] * *vé* au lieu de « veux » (cf. [4.a ; 5.b])

1.2. Principes de création de mots propres au FPI

Le FPI utilise d'autres procédés de création de mots nouveaux ; ces principes vont de la transformation d'une nouvelle consonne à l'onomatopée en passant par la troncation et à l'emprunt.

– *La nasalisation*

Sous l'influence du substrat linguistique des langues maternelles africaines, chez les locuteurs du FPI, la consonne orale dento-latérale [l] du mot français, se transforme automatiquement en une consonne nasale dento-alvéolaire [ɲ] dans un contexte phonologique où, d'une part apparaissent le trait de la nasalité et d'autre part la consonne [n] : il s'agit d'un cas d'assimilation régressive totale de la consonne orale [l] sous l'effet de la consonne nasale [n] ou même du trait de la nasalité ; ainsi,

[balɔ̃] « ballon » devient-il * [banɔ̃] « banon » (cf. 11)

[elefã] « éléphant » * [nedefã] « nénéphant » (cf. 9).

– *La liaison obligatoire*

Il n'existe pas de déterminant article antéposé au substantif dans les langues ivoiriennes si bien que, lorsqu'en français la liaison devient obligatoire avec un mot commençant par une voyelle, le locuteur du FPI saisit les deux unités du syntagme nominal comme une seule entité ; ainsi, un éléphant deviendrait *néléphant. Or les langues ivoiriennes n'acceptent pas pareille structure, la dentale nasale [n] et orale [l] ne pouvant figurer dans un mot : car dans ce cas, la latérale orale est assimilée par la nasale ; on obtient alors en FPI le mot :

* [nedefã] « nénéphant » au lieu de [elefã] « éléphant » (cf. 9)

– *Troncation*

Le FPI crée également des mots par troncation du mot français ; cette opération s'effectue généralement à la fin du mot français et concerne la consonne finale d'une syllabe fermée ; par exemple :

[brezil] « Brésil » devient-il en FPI : * [brezi] « Bresi » (cf. 11)

[avɛk] « avec » * [avɛ] « avai » (cf. 12)

– *Onomatopée*

Le FPI forme des mots dont le phonétisme suggère par duplication vocalique une certaine harmonie imitative ; ils s'apparentent de ce fait aux onomatopées classiques obtenues par duplication syllabique ou lexicale. Ces mots, qui souvent ont valeur d'interjections, portent une charge affective intense traduite par l'allongement démesuré de la syllabe ou voyelle finale, allongement d'ailleurs appuyé par une double ponctuation ; par exemple :

héi ! ? « hé » (cf. 5.a)

hééé !!! « hé » (cf. 6.a)

dépuiiii... « depuis » (cf. 6.b)

– *Emprunts*

Le FPI se veut une langue complète, qui entend couvrir la totalité de l'expérience humaine en enrichissant son lexique. Ainsi, quand cette langue ne trouve pas de mot disponible approprié en français pour exprimer une réalité de son milieu de vie, elle a systématiquement recours à l'emprunt : elle puise ainsi dans le lexique des langues ivoiriennes les mots qui lui sont indispensables à exprimer sa vision du monde et sa sensibilité ; par exemple :

yako ! (cf. 11)

djigbô (cf. 12)

Ces emprunts figurent en bonne place dans les dictionnaires du FPI, tel celui de Lafage ; ils y sont définis comme suit :

yako : « Expression exprimant la tristesse et signifiant que l'on compatit à la douleur de quelqu'un. Hélas ! Condoléances ! Présenter ses condoléances, exprimer sa compassion. »³

djigô : « Fétiche à influence maléfique. Recourir à la magie, pratiquer un envoûtement (contre qqn) »⁴

Certes, le *djigô* est un fétiche mais il n'est pas forcément maléfique. Dans le *corpus* que nous exploitons, gagner un match de football ne peut être que bénéfique aux supporters de cette équipe ; d'autre part, si le mot *yako* conserve le sens général défini dans le dictionnaire, il est à noter que dans notre *corpus*, *yako* ! est plutôt employé dans un sens ironique par Moussa qui plaint l'étroitesse d'esprit de son ami Zézé, incapable de comprendre que l'équipe de football du Brésil en 1986, est en baisse de régime et ne peut donc pas remporter la coupe du monde.

1.3. Modèles de mots absents du *corpus*

Nous parlerons également de certains modes d'enrichissement lexical en FPI dont les traces n'apparaissent pas dans notre *corpus*, trop restreint : il s'agit de mots français qui acquièrent des sens nouveaux en FPI ou des mots qui sont la marque d'un produit industriel ou encore des mots qui désignent le pays de provenance d'un produit commercial et qui, par dérivation impropre, deviennent des substantifs.

– *Mots français de sens nouveaux en FPI*

Des mots français passent en l'état morphologique en FPI, mais ils y sont enrichis de sens nouveaux jusque-là ignorés en français standard ; il en est ainsi des mots comme : *maquis*, *mon mari m'a laissée*, *deuxième bureau*...

Le mot le plus prospère de tous est *maquis* ; il relève du secteur économique de l'informel et est de ce fait quotidiennement utilisé par une frange importante de la po-

³ Suzanne Lafage : *Le Lexique français de Côte-d'Ivoire*, t. II, pp. 823.

⁴ Suzanne Lafage : *op. cit.*, t. I, pp. 321.

pulation. Le mot serait « africanisé » puisqu'on le retrouve dans le dictionnaire de l'équipe de l'IFA (2004) et dans celui de Lafage (2002) où il apparaît, en Côte d'Ivoire, dans plusieurs expressions idiomatiques relevées dans les journaux ou chez les romanciers ivoiriens. En français standard, *maquis* désigne un :

« Paysage méditerranéen dont la végétation dense et touffue est composée de petits arbustes, de bruyères, de lauriers, poussant sur des terrains siliceux.

Lieu retiré, en général boisé ou montagneux, dans lequel se réfugiaient et luttaient les résistants, pendant l'occupation allemande en France, de 1940 à 1944. »⁵

L'Équipe de l'IFA fait apparaître les différentes assertions du mot dans chaque contrée africaine :

« En Côte d'Ivoire : Restaurant semi-clandestin où l'on consomme surtout du gibier (fourni par les braconniers) ; au Sénégal : Bar, dancing, hôtel fréquenté par les prostituées et les mauvais garçons ; à Madagascar et au Sénégal : Ensemble des lieux mal famés d'une ville. »⁶

Dans *Le Lexique français de Côte d'Ivoire*, Lafage le définit comme suit :

« Usuel, oral, écrit, tous milieux, autrefois péjoratif, le terme est devenu mélioratif et synonyme de lieu agréable et convivial.

Restaurant populaire où l'on consomme des plats africains et parfois du gibier braconné. Il sert aussi de bar, de dancing, parfois même d'hôtel... »⁷

Les maquis, ces restaurants à bon marché et à excellente nourriture typique... qui vibrent au rythme du zouk et du funk jusqu'à une heure avancée, quel que soit le jour de la semaine... seraient nés, pense-t-on, d'initiatives ghanéennes.

Les maquis vendraient d'abord en fraude de la viande de brousse et du gin artisanal, d'où leur nom ; pour les uns encore, cette appellation viendrait du fait que les maquis servaient de lieux de rencontre où les habitués – appelés “maquisards” – parlaient librement politique et refaisaient ainsi le monde [au temps du parti unique et de la pensée unique, avant 1990] ; pour d'autres encore, le nom viendrait de leur illégalité, ces restaurants de fortune ne disposant d'aucune autorisation pour exercer leur art. »⁸

– *Mots de marque de fabrique*

On retrouve aussi dans le FPI des mots usuels qui ne sont rien d'autres que des marques de produit ou tout simplement des noms d'usine de fabrication de ces produits ; par exemple : *macaci, mipa, fancy, wax...*

⁵ Jean Dubois : *Lexis du français*, cf. maquis.

⁶ Équipe IFA : *Inventaire des particularités lexicales du français en Afrique Noire*, cf. maquis.

⁷ Suzanne Lafage: *op. cit.*, t. I, pp. 558-559.

⁸ Suzanne Lafage: *op. cit.*, t. I, pp. 558-559.

Mipa est une usine spécialisée dans la fabrication d'ustensiles de cuisine tels les cuvettes, les verres, les assiettes... Cette usine utilise le caoutchouc comme matière première. En FPI, *mipa* sert à désigner un article ménager fait en plastique et qui occupe une place importante dans la vie quotidienne de la ménagère ivoirienne ; *mipa* s'oppose de ce fait à tout ustensile de cuisine en émail, en verre ou en porcelaine qu'utilisent les personnes nanties.

De même, *fancy* désigne un tissu localement produit par l'usine de fabrication de pagnes de Côte d'Ivoire, dénommée Fancy.

– Mots désignant un pays d'origine

Des mots désignant le pays d'origine d'un produit commercial populaire en Côte d'Ivoire entrent désormais dans le lexique du vocabulaire du FPI ; ces mots sont souvent des substantifs adjectivaux tels que : *hollandais*, *thaïlandais*...

Hollandais désigne un pagne en provenance de la Hollande, pays spécialisé en la matière et qui dessert le marché ivoirien en pagnes de luxe ; le *hollandais* s'oppose au *fancy*, de fabrication ivoirienne, de moindre qualité et par conséquent de moindre coût.

Le *thaïlandais* ne réfère nullement à un ressortissant de la Thaïlande mais à une espèce de riz prisé en Côte d'Ivoire et importé de ce pays. Malgré la qualité de ce riz, il est vendu à un prix concurrentiel sur le marché ivoirien.

2. LE NOUCHI

Le nouchi est la dernière variété de français ivoirien née dans les années quatre-vingt. Comment est née cette langue ? Quel est son mode de création de mots ?

2.1. Origine du nouchi

L'origine du nouchi est à décrypter à travers la morphologie du mot. Étymologiquement, « nouchi » est un mot mandingue ; morphologiquement, il est formé de la juxtaposition des deux monèmes suivants : « *nou* » qui signifie « narine » et « *chi* » qui signifie « poils ». Littéralement traduit, nouchi désignerait donc « les poils qui débordent des narines ». Ainsi dit, « *nouchi* » devient un symbole qui fait référence aux jeunes délinquants et enfants de la rue qui vivent dans un état de dénuement total au point de ne pouvoir prendre soin de leur corps. En effet, le phénomène des enfants de la rue est inconnu en Côte d'Ivoire avant les années 1980. Il apparaît avec les vagues de jeunes élèves issus de « l'enseignement télévisuel » et qui sont massivement renvoyés du collège.

Ces enfants, encore trop jeunes pour être embauchés dans une usine, vont se retrouver dans la rue sans que l'état sache intervenir avec promptitude pour résoudre leur problème de survie. Ces enfants, exclus du collège dès le cycle d'observation, en classe de 6^e et 5^e, sont tous nantis du certificat d'études primaires élémentaires (CEPE). Ils ont donc appris le français standard à l'école et le pratiquent couramment, du moins à l'oral. Cependant, enfants de la rue, ils ont préféré créer une langue à eux, le nouchi,

inaccessible aux adultes et qui leur permette de se passer des consignes afin d'opérer des vols à la tire. Ainsi, le nouchi serait un argot de petits voyous. Mais aujourd'hui, tous les jeunes citadins, élèves, collégiens et étudiants se sont appropriés cette langue en réaction contre la suprématie du français en Côte d'Ivoire. La langue des chanteurs de variétés musicales ivoiriennes, créateurs du « zouglou » qui sont pour la plupart des étudiants, n'est-elle pas le nouchi ? Aussi Kouadio définit le nouchi comme suit :

« Né dans la rue, ce parler est devenu le code de ralliement d'une majorité des jeunes Ivoiriens : élèves, lycéens, étudiants, jeunes de la rue, jeunes délinquants. »⁹

2.2. Création lexicale en nouchi

Le lexique du *corpus* nouchi suivant reflète parfaitement le niveau culturel de ses locuteurs :

Saïdou : – « Les gus me voilà !!

On peut faire la bringue jusqu'à demain ! »

Les copains de Saïdou : – « Saïdou, dépêche-toi !

Les dayes nous attendent au zigli-bity-club »¹⁰

Le *corpus* ci-dessus est bien évidemment un dialogue entre Saïdou, élève en classe d'examen que ses parents croient s'être enfermé dans sa chambre pour faire ses révisions mais qui en réalité s'est échappé par la fenêtre arrière afin de rejoindre des amis avec qui il a prévu d'aller au dancing, au Zigli-bity-club. On relève de leur langage des mots de l'argot français mais aussi des mots de composition anglaise ou d'association de monèmes des langues ivoiriennes et du français.

– Mots de l'argot français

On note chez le locuteur nouchi des mots argotiques rarement utilisés en français standard en Côte d'Ivoire tels que « gus » et « bringue », ce qui dénote de leur degré de maîtrise du français standard, langue d'enseignement, apprise et parlée en classe. Nous savons qu'en français *gus* est synonyme de « type, mec » et que *bringue* signifie « faire bombance ».

– Mot des langues étrangères européennes

Le nouchi emploie des mots créés à partir d'autres langues européennes, notamment l'anglais et l'espagnol appris au collège ; par exemple le mot *daye*. Dans le dictionnaire de Lafage, il est expliqué comme suit :

« daye, n. f. Argot étudiantin et urbain, (anglais "mourir") oral, fam, péj. ivre mort, complètement saoul. »¹¹

⁹ J. N'G. Kouadio : *Le Nouchi abidjanais; naissance d'un argot ou mode linguistique passagère ?*, p. V.

¹⁰ *Ivoire-Dimanche*, magazine de Côte d'Ivoire, spécial *Fraternité-Matin*, n° 808 du 3 août 1986, pp. 50.

Parti de « (to) die », mot anglais signifiant « mourir », le mot « daye » par extension veut dire en nouchi « être amoureux fou d'une fille comme si l'on avait été envoûté par elle ».

Les étudiants emploient aussi des mots espagnols pour désigner leurs géniteurs ; ils disent : « mon padré », « ma madré » : *padre* et *madre* ; ce sont des mots espagnols qui sont précédés d'un déterminant article français et qui sont affectés d'accents aigus, car en espagnol, le « e » se prononce comme le « é » français et par conséquent, il n'a pas besoin d'une marque accentuelle.

– *Mot ivoirien et mot français*

Le nouchi crée également des mots nouveaux en alliant un monème des langues ivoiriennes et un monème du français ; ainsi est formé le mot « ziglibity-club », appellation d'un dancing d'Abidjan. En effet, le « ziglibity » est mot ivoirien ; il désigne une variété de danse traditionnelle ivoirienne qui a été modernisée par le musicien Ernesto Djédjé. Celui-ci se produisait dans son « *ziglibity-club* » où se retrouvaient les adeptes de sa musique pour se divertir. Mais on pourrait aussi citer un autre mot qu'affectionnent les locuteurs du nouchi : « *bras mogo* » ; il s'agit ici d'un mot composé, du mot français *bras* et d'un mot jula « *mɔ`gɔ´* » qui signifie « homme ». *Bras mogo* signifie originellement en nouchi : « ouvrier ; manutentionnaire » ; avec le temps, il y a eu glissement de sens et cette expression signifie aujourd'hui un camarade galérien, un ami.

Les variétés des français ivoiriens sont donc caractérisées par leurs procédés de création lexicale mais surtout par les populations qui sont à la base de leur création ou qui majoritairement les parlent ; ainsi, le FPI est pratiqué par des populations ivoiriennes et africaines analphabètes vivant en milieu urbain et rural tandis que le nouchi, qui se signale par sa vitalité comme la future langue ivoirienne, demeure pour l'instant un phénomène urbain propre à une jeunesse déscolarisée mais aussi estudiantine en mal d'une quête identitaire. Ainsi, « À tort ou à raison, écrivions-nous en 1989, la Côte d'Ivoire s'est taillé une certaine réputation pour son français. Non pas certes pour le français soutenu des élites lettrées ou universitaires [...], mais pour son français populaire parlé par les locuteurs moyens, peu ou pas lettrés »¹².

3. TRADUCTION ET CONTEXTE CULTUREL

Du fait que les variétés ivoiriennes du français sont exclues du cadre institutionnel pour être confinées à la clandestinité, l'étude exhaustive d'une langue comme le nouchi n'est pas encore faite alors que cette langue mériterait d'être mieux connue sur le plan scientifique, par exemple, des enseignants dans le but d'une interprétation rationnelle des écarts de langue écrite ou orale rencontrés dans les productions des élèves et éco-

¹¹ Suzanne Lafage : *op. cit.*, t.1, pp. 277.

¹² B. Mel et J. N'G. Kouadio : *Visages du français variétés lexicales de l'espace francophone : cas de la Côte d'Ivoire*, p. 5.

liers ivoiriens. Une étude scientifique du nouchi préparerait cette langue à une fonction noble telle son introduction dans l'enseignement suite à son enrichissement lexical.

Mais l'absence d'une connaissance exhaustive et scientifique des variétés ivoiriennes du français n'empêche pas que des dictionnaires soient consacrés aux particularités lexicales de ces français. On retrouve ainsi dans ces dictionnaires des mots directement puisés dans les langues ivoiriennes ou des mots de morphologie française mais dont les sens, dans le contexte ivoirien, empruntent des valeurs que ne saurait correctement exprimer la seule métalangue du dictionnaire classique, si bien qu'il s'avère nécessaire d'y adjoindre image et son pour rendre intelligible à tout francophone non ivoirien les sens que véhiculent ces termes spécifiques. Nous traiterons ici deux de ces termes à titre d'exemples : les mots *maquis* et *dipri*.

3.1. Mots français en contexte ivoirien

Les dictionnaires français nous donnent une définition exacte du mot *maquis* en français standard. La signification de ce mot en contexte ivoirien nous est également proposée par les dictionnaires des variétés du français africain et ivoirien. Aussi note-t-on qu'en FPI et dans l'usage ivoirien du mot, la notion de paysage disparaît de la perception du mot *maquis* ; on ne retiendra comme assertion qu'un lieu de refuge pour personnes en rupture de ban avec la société. De nos jours, le *maquis* est davantage un milieu africain de restauration, de boisson et de musique, c'est-à-dire un type d'exploitation économique citadin qui se répand avec la forte récession que subit le pays et un mode de reconversion économique des travailleurs massivement licenciés des entreprises. Disons que le *maquis* est devenu aujourd'hui un mode de vie dans les grandes villes de Côte d'Ivoire et notamment à Abidjan où une rue leur est consacrée : « la rue princesse ». Cette rue est devenue de ce fait une curiosité touristique dont les charmes nocturnes n'échappent à aucun étranger visitant Abidjan.

Ces différents sens qu'emprunte le mot en FPI, ainsi que les activités économiques et culturelles qu'il implique, nous éloignent du sens originel du mot en FS.

D'ailleurs il en est ainsi d'une expression idiomatique FPI qui présente une coloration française mais dont le sens échapperait à tout francophone ; il s'agit de l'expression « mon mari m'a laissée ». Tout naturellement, cette phrase nous fait penser à une femme répudiée ou à une veuve. En FPI, même si elle peut avoir pour auteur une femme, le référent de cette phrase est loin d'être un être animé, à plus forte raison une personne humaine ; car à Abidjan, *Mon mari m'a laissée* est le nom d'un quartier, quartier célèbre par l'état de misère qu'il inspire et qui contraste violemment avec la prestance des personnes qui y vivent. Il s'agit généralement d'infortunées concubines analphabètes qui ont servi de « deuxième bureau » (ou de maîtresses) aux hauts cadres ivoiriens. Elles ont été habituées au luxe, et l'éclat de leur peau ou le style de leur habillement le prouve encore ; mais sous l'effet de la crise économique qui frappe durement le pays, ou découvertes par l'épouse légitime, ces pauvres dames ont été abandonnées un matin par leurs concubins, parfois avec des enfants illégitimes à leur charge. Impuissantes face aux lois du pays qui imposent la monogamie, démunies de tout moyen, elles n'ont dû leur salut qu'au fait d'aller se réfugier dans un lieu insalubre, un quartier

précaire où les habitations sont construites avec des matériaux de récupération tels que des feuilles en plastique, de vieilles tôles. *Mon mari m'a laissée* est donc avant tout un cri de détresse de femmes déshéritées, malheureuses, qui hurlent leur douleur dans la pauvreté ; l'humour ivoirien qui a tendance à tourner en dérision le malheur et à le sublimer pour mieux le surpasser n'est pas absent de cette expression.

3.2. Des mots africains dans le dictionnaire du FPI

À titre d'exemple des mots africains passés dans le dictionnaire du FPI, nous citerons le mot *Dipri*. Le mot est défini dans le *Lexique français de Côte d'Ivoire : appropriation et créativité* de Lafage comme suit :

« Fête célèbre pour ses rites de possession, ses transes, les blessures au couteau que s'infligent les *séképouéné* (les possédés) apparemment sans douleur ni cicatrices visibles par la suite, le *Dipri* commémore une terrible famine chez les Abidji, à laquelle mit fin le sacrifice du fils du chef Bidyo, coupé en morceaux et enterré. Des ignames poussèrent spontanément sur le charnier. »¹³

Cette définition est tout à fait juste. Mais comment faire accéder par la seule puissance du mot du dictionnaire la réalité de la fête du *Dipri* à un étranger ? Car, culte solennel rendu aux ancêtres à chaque année nouvelle, à la fin du mois d'avril, le *Dipri* est l'occasion à la fois de remercier les divinités et les ancêtres divinités, les *awenté*, de leurs bienfaits durant l'année écoulée, mais aussi l'occasion de solliciter leur protection pour la nouvelle année et d'honorer les défunts récents. Or, invoquer les divinités suppose que l'on soit en règle avec elles, que l'on répare les transgressions dont on s'est rendu coupable vis-à-vis des membres de la société, si bien que le *Dipri* apparaît également comme une fête de réconciliation, d'où la présence impérative au *Dipri* de tous les membres de la tribu.

« Très vivace donc dans les villages Ogbu, elle se déroule avec un particulier éclat à Yaobou, Sahouyé et surtout à Gomon où elle est rehaussée par les manifestations magiques *sékè*. »¹⁴

En effet, lors du *Dipri*, les *Eikpa*, personnes dotées de certaines puissances naturelles, renouvellent ou accroissent leurs forces ou *sékè* auprès des génies de la rivière de Gomon : dès lors, elles se livrent à des démonstrations publiques de leurs puissances magiques ; ainsi, l'ethnologue Lafargue qui, à plusieurs reprises, a assisté au *Dipri* en témoigne dans les termes suivants :

« Certains *kponpuoné* faisaient sortir leurs yeux des orbites et les faisaient pendre sur leurs joues, puis les remettaient en place ; d'autres coupaient un enfant en morceaux le matin et le reconstituaient vivant le soir »¹⁵

¹³ Suzanne Lafage : *op. cit.*, t. 1, pp. 315.

¹⁴ Fernand Lafargue : *Sorcellerie*, pp.222.

¹⁵ Fernand Lafargue : *op. cit.*, pp.235.

Il persiste et signe sa déclaration en ces termes :

« Nous avons observé nous-même ces faits que nous venons de relater. »¹⁶

Comment un dictionnaire peut-il restituer ces scènes poignantes ? Nous croyons qu'il faut aujourd'hui associer mots-sons-images, c'est-à-dire les scènes de vie des populations et cultures des peuples de la Francophonie sur un support audiovisuel pour rendre vivante, palpable et intelligible une réalité étrangère aux observateurs issus d'une autre civilisation. Du reste, ce procédé nous permettra de mieux faire connaître et mieux faire apprécier les divers peuples de la Francophonie entre eux.

La problématique qui a été traitée ici est celle de la présentation ou de l'enseignement du signe linguistique dans un contexte étranger : peut-on se servir de simples mots pour parler par exemple de la neige à un écolier ivoirien vivant sous les tropiques ? L'association de l'image, de la scène de vie courante et du son ne sont-ils pas nécessaires pour faciliter à l'étranger la tâche de la compréhension de la réalité culturelle d'une autre société ?

¹⁶ Fernand Lafargue : *op. cit.*, pp.235.

CORPUS FPI

- Bonjour, moi c'est Yao. Et toi ?
- Dago – On me pelle Dago ¹ *On m'appelle Dago*
- Moi, y'en a content aujourd'hui. ¹ *Moi, [il] y en a la joie aujourd'hui
= je suis content*
- Moussa – Zézé, ti a vi !! *Zézé, tu sais,*
- Mon pétit est sérié ¹ *Mon garçon [que voici] est [très]
sérieux*
- Zézé – Yé vé tousse mon seque-là ! *Je veux toucher mon chèque !*
- Heï ! ? Pendant un moi zé m'a fatigue *Hé ! [voici] un mois [que], je me
pour gangné l'arzen-là !* *fatigue à avoir mon argent !*
- Auzoud'hui yé suis véni sarse ça, wala on *Aujourd'hui [que] je me suis décidé
me fatigue encore ! ?* *à venir chercher ça, voilà qu'on me
créé encore des ennuis !*
- Ça vé dit quoi ça ? ¹ *Quelles sont ces manières-là ?*
- Zézé – Héeé ! Dié !! Si comme zai gagné l'arzent, *Bon Dieu ! Si j'étais riche depuis
dépiiii yé suis messico là-bas !* *longtemps je serais parti au Mexico !*
- Moussa – Messico ??... Messico où la tair tue *Au Mexique ? Au Mexique où la
n'homme !* *terre tue les gens, c'est là-bas que tu
veux aller ?*
- Zézé – Toi-aussi, trembément dé tair c'est pas *Écoute, les tremblements de terre
touzour !! E puis si comme yé suis là-bas yé pé* *n'arrivent pas tous les jours !*
- courazé mon 'quipe vis-à-vis ! *Si donc j'étais là-bas, j'allais en-
courager mon équipe de vive voix !*
- Moussa – Hi ! Hi ! Hi ! Toi-même tu vas couraze *Beauf ! Qui vas-tu encourager ?
qui ça ?* *Est-ce les Éléphants se sont rendus là-
bas ?*
- Asque Nénéphant i l'a voyazé zousqu' àà là-
bas ?
- Zézé – Hooo ! Quitte-là !! Touzou dans Moundia *Ho ! Fous-moi la paix ! Toujours
c'est Brésil que je supporte ! Tu sais ça !* *au mondial, c'est le Brésil que je sup-
porte ; aurais-tu oublié ça ?*
- Moussa – Alors Zézé, yako ! Auzoud'hui-là, banon *Alors Zézé, du courage ! Ne sais-tu
de Brézi c'est pétit !* *pas qu'aujourd'hui, le football du
Brésil est en forte baisse ?*
- Zézé – Avai mon djigbô niveau-model-là Bassam *Ne sais-tu pas qu'avec mon nou-
lui-même pé véni Sampion de Moundia 86 ?¹* *veau fétiche, même une équipe de
locale comme celle de Bassam serait
championne au Mondial 86 ?*

BIBLIOGRAPHIE

- Dubois, (J.). *et al*, 1980 : *Lexis du français*, Paris, Larousse.
- Équipe IFA, 2004 : *Inventaire des particularités lexicales du français en Afrique Noire*, Paris, EDICEF/AUF, 443 pp.
- Ivoire-Dimanche*, magazine de Côte d'Ivoire, spécial *Fraternité-Matin*, n^{os} 167 et 168 d'avril 1974 ; n^{os} 784 du 16 fév. et 808 du 3 août 1986, Abidjan, Fraternité-Matin.
- Kouadio, (J. N'G), 1990 : « Le Nouchi Abidjanais ; naissance d'un argot ou mode linguistique passagère ? » dans *Actes du colloque des langues et des villes*, Dakar, 15-17 Didier Érudition.
- Lafage, (S.), 2002 : *Le lexique français de Côte d'Ivoire. Appropriation et créativité*, Nice, t. I et t. II, 861 pp.
- Lafargue, (F.), 1976 : *Sorcellerie des Abidji en Côte d'Ivoire*, Paris, Nouvelles éditions latines, 302 pp.

VARIABILITÉ ET VARIATION EN TERMINOLOGIE ET LANGUES SPÉCIALISÉES : DISCOURS, TEXTES ET CONTEXTES

Isabel Desmet

département de portugais
Université Paris VIII
Saint-Denis, France

Résumé : Dans la présente communication, nous proposons de (re)définir variabilité et variation dans les terminologies et langues spécialisées. La variation des terminologies étant directement tributaire de la variation textuelle et discursive, nous proposons également des typologies de discours, de textes et de contextes de spécialité. Enfin, la variation discursive, textuelle et contextuelle nous conduit à proposer un modèle d'analyse des termes où ceux-ci sont perçus comme des unités lexicales spécialisées, ayant des sens, des emplois et des usages particuliers.

Mots-clés : variabilité, variation, terme, discours, texte, contexte

1. INTRODUCTION

La linguistique, qui a pour objectif fondamental la description des langues, se confronte naturellement à l'étude de la variation qui les caractérise. Il est également naturel que l'étude linguistique des langues de spécialité – et de leurs unités d'analyse, les termes — intègre une optique variationniste, car les langues de spécialité sont, avant tout, des langues naturelles dans leur fonction de communication de savoirs spécialisés (*cf.* Lerat 1995). Or, un tel programme n'est que depuis peu préconisé par des linguistes des langues de spécialité et, le plus souvent, seulement de manière partielle. Ceci s'explique par l'histoire même des études sur les langues de spécialité, en particulier par l'histoire des études terminologiques, qui ne trouvent pas leurs origines dans la linguistique descriptive.

Les emplois des mots spécialisés dépendent souvent des types de textes et de discours. L'étude des plans supraphrastiques – textuel et discursif – doit accompagner toute étude des différents niveaux d'analyse linguistique, du morphématique à l'énonciatif. Le temps et le lieu de production d'un texte donné doivent aussi faire partie intégrante de toute analyse linguistique des langues spécialisées. En somme, il est question ici d'une pragmatique intégrée et non d'une pragmatique *a posteriori*. Seule une approche de ce type permettra de saisir les différents types de variations inhérents aux langues de spécialité.

2. TERMINOLOGIE ET LANGUES SPÉCIALISÉES : VARIABILITÉ ET VARIATION

Nous définissons la « variabilité » comme la capacité de toute langue naturelle de produire de la variation lorsqu'elle s'actualise en discours. La variation, sous toutes ses formes et à tous les niveaux ou plans d'analyse linguistique, est en quelque sorte la conséquence directe de la variabilité, inhérente à toute langue naturelle. De notre point de vue et d'après notre pratique, les termes et les langues spécialisées n'échappent pas à ce phénomène naturel, tout comme les mots de la langue générale et la langue générale elle-même. Ce qui aux yeux d'un linguiste peut paraître une évidence ne l'est pas pour une bonne partie des terminologues. En fait, selon la théorie générale de la terminologie, le terme est perçu uniquement comme une dénomination, non variable, normalisé. Il est attaché à un domaine et la seule perspective synchronique intéresse les études terminologiques. Le poids de cette tradition se reflète encore dans des écrits récents. Dans d'autres écrits également récents, on commence cependant à voir la variation en terminologie et dans les langues spécialisées prises en compte, ce qui fait même de cette problématique un sujet central des derniers temps. L'étude de la variation dans les terminologies et dans les langues de spécialité représente un tournant radical, théorique et méthodologique, qui a des conséquences directes dans les différentes applications des études terminologiques.

Nous proposons une approche des langues de spécialité fondée pour l'essentiel sur les principes de la linguistique descriptive, intégrant la variation à plusieurs niveaux : variation des langues, des discours et des textes de spécialité ; variation dans le temps, dans l'espace, en fonction des interlocuteurs et des situations de communication ; variation des unités linguistiques de spécialité sur l'axe syntagmatique et sur l'axe paradigmatique ; variation sur le plan lexical, sur le plan phrastique, sur le plan textuel et sur le plan discursif. À un modèle de traitement des unités terminologiques qui privilégie le lexique par rapport à la grammaire, au nom de la priorité et précision des concepts (modèle standard de la terminologie), nous opposons un modèle comportant les niveaux d'analyse typiques de la linguistique : lexical, syntaxique, sémantique et pragmatique. Le niveau pragmatique commence par la prise en considération de la variation discursive, textuelle et contextuelle, qui se manifeste aux niveaux phrastique, sémantique, lexical, morphologique et même phonétique. Ainsi une focalisation sur les discours, les textes, les contextes et les termes devient nécessaire.

3. VARIATION DISCURSIVE : TYPES DE DISCOURS DE SPÉCIALITÉ

Le terme « discours », on le sait, recouvre différentes acceptions selon le point de vue théorique qui le fonde comme concept, et le terme « analyse » correspond à des procédures différentes selon la théorie qui le définit.

Indépendamment des différentes positions théoriques, la première grande distinction concerne les spécialités en soi : le discours de la médecine, du droit, des sciences naturelles, etc. Si l'on admet que les discours spécialisés reflètent les caractéristiques conceptuelles des différents domaines du savoir, il convient alors de procéder à des

distinctions claires par rapport à la diversité des systèmes conceptuels (cf. Rey 1979 : 44).

Si l'on admet encore que les langues spécialisées reflètent les caractéristiques notionnelles des différents domaines du savoir, il faut également prendre en compte le fait que les langues spécialisées se caractérisent par des niveaux différents par rapport à leurs contenus. L. Hoffmann (1976 : 184-192) et R. Kocourek (1991 : 37) découpent la langue de spécialité d'après quatre critères : le degré d'abstraction, la façon naturelle/artificielle d'exprimer les éléments et la syntaxe, le milieu ou type de spécialité et les participants. L'application de chaque critère mène à une stratification de la langue en cinq niveaux. Le niveau le plus élevé est celui du langage symbolique et ne fait partie de la langue technoscientifique que par sa composante en langue naturelle. Le niveau le moins élevé est celui qui correspond à la langue de consommation assez proche de la langue usuelle. Ce classement classique nous semble toujours important dans la mesure où il vient éclairer le problème de la diversité discursive.

Un certain nombre de catégories de discours scientifique ont été identifiées notamment par A. M. Laurian (1983 : 10-12), J. Pearson (1998 : 35-39) et I. Meyer et K. Mackintosh (1996 : 270-271). Cette catégorisation, essentiellement basée sur la situation de communication, fait que les discours diffèrent entre eux selon l'émetteur du discours, le récepteur ou destinataire du discours, le but du discours et le degré de technicité ou de spécialisation du discours.

Les catégories que l'on retrouve de manière systématique chez tous les auteurs peuvent être schématisées de la façon suivante :

<i>Discours de spécialité</i>
- Discours scientifique spécialisé (p. ex : un manuel spécialisé ; une revue scientifique)
- Discours scientifique officiel (p. ex : des textes de loi)
- Discours scientifique pédagogique ou didactique (p. ex : un manuel pour les étudiants)
- Discours de semi-vulgarisation scientifique (p. ex : une revue de spécialité pour un public d'initiés)
- Discours de vulgarisation scientifique (p. ex : les secteurs spécialisés dans la presse générale)

Tableau 1 : Types de discours de spécialité

En somme, lorsqu'il s'agit de traiter les langues de spécialité, leurs textes peuvent être choisis en fonction de leur appartenance à un genre textuel ou à des genres textuels et à un type ou à des types de discours scientifique et technique, qui correspondent à la traditionnelle division verticale des langues de spécialité.

4. VARIATION TEXTUELLE : TYPES DE TEXTES SPÉCIALISÉS

D'un point de vue strictement linguistique, la linguistique textuelle se distingue des autres disciplines d'interprétation qui constituent le champ multidisciplinaire de l'analyse de discours, bien qu'elle en soit connexe et qu'elle y trouve des éléments communs d'analyse. Dans cette perspective, les textes sont, dans un premier plan, la source des données pour l'analyse de tous les plans de la langue.

L'ensemble des textes spécialisés peut être défini comme la forme, parlée et écrite, et le contenu de tout ce que l'on exprime dans les langues spécialisées. Cette définition, certes trop vaste, présente l'avantage de prendre en considération les textes oraux et les textes écrits, les textes contemporains et les textes passés, des textes en langue standard et non standard, des textes en langue centrale et des textes en langue régionale, des textes primaires ou hautement spécialisés et des textes de vulgarisation et didactiques...

Les textes scientifiques et techniques prennent la forme de l'un des types habituels qui sont donnés par la spécialité et par sa diversité, par le sujet dont on traite, par la tradition ou des habitudes stylistiques, et par leurs objectifs et fonctions (ensemble de critères distinctifs portant sur le contenu thématique et découpage en domaines, contenu circonstancié et niveaux d'abstraction ou division horizontale et division verticale, niveaux scientifiques et niveaux didactiques). Ces différents types se constituent en fonction des besoins de l'activité quotidienne des spécialistes : étude, manuel, thèse, monographie, mémoire, article de périodique, dissertation, exposé, communication, etc. (*cf.* Kocourek, 1991 : 48).

Les critères de classification des textes spécialisés varient selon l'objectif. Outre les critères déjà mentionnés, on peut encore établir des distinctions entre un texte continu et une liste (un manuel et un catalogue), entre un texte simple (un résumé, par exemple) et un texte complexe (un manuel, une monographie ou une thèse), des textes fondamentaux et des textes utilitaires... Enfin, toutes les formes de variation de la langue et des langues s'appliquent également aux textes (variation diachronique, diastratique et diatopique).

L'étude des différents plans d'analyse linguistique d'un ensemble de textes spécialisés a conduit certains auteurs à élaborer des typologies partielles. C'est le cas, par exemple, des caractéristiques définitives d'A.-M. Laurian (1983 : 12-20). Cependant, les études les plus complètes de typologies textuelles proviennent de la linguistique textuelle générale. Les travaux de R. Gläser (1993), G. Weise (1993) et W. Heinemann (1991 et 2000) proposent des typologies complexes ou modulaires, à fondements cognitifs et communicatifs, axées sur différents niveaux ou paramètres textuels. Selon les travaux de W. Heinemann (2000), nous pouvons distinguer quatre niveaux dans les textes, en général : le niveau fonctionnel, le niveau situationnel, le niveau du contenu sémantique et le niveau formel-grammatical. Cette typologie, qui représente une bonne base de travail, peut être résumée comme suit :

<ul style="list-style-type: none"> – Fonctions : – exprimer – contacter – informer – diriger – Hierarchie fonctionnelle (structure illocutive) : – fonctions dominantes – fonctions subsidiaires – fonctions complémentaires – Séquences fonctionnelles 	<ul style="list-style-type: none"> – Communication interne ou externe à la discipline : – interne (état, religion, commerce, industrie, etc.) – externe (communication interdisciplinaire, communication avec le monde public) – Interlocuteurs : spécialiste-spécialiste ; spécialiste-initié spécialiste-grand public – Relation entre les interlocuteurs – nombre d’interlocuteurs (monologue, dialogue) – Paramètres spatio-temporels – (communication présente, graphique, visuelle, etc.) 	<ul style="list-style-type: none"> – Sujet du texte – Perspective sur le sujet (théorique, didactique, appliquée, de divulgation, etc.) – Parties textuelles (libres ; standardisées) – Développement thématique (séquences descriptives, narratives, expositives, argumentatives, directives) 	<ul style="list-style-type: none"> – Maximes rhétoriques – Formes linguistiques et non-linguistiques – Aspects lexicaux et grammaticaux (ressources lexicales et syntaxiques)
--	--	--	---

Tableau 2 : Typologie pluri-niveaux de textes de spécialité

Le niveau fonctionnel est celui des fonctions textuelles : s’exprimer, contacter, informer, diriger. Les textes peuvent être monofonctionnels ou plurifonctionnels, ce qui implique le besoin d’analyser la hiérarchie des fonctions dans un texte (fonctions dominantes, subsidiaires et complémentaires), ainsi que le besoin d’identifier les séquences des fonctions dans les textes.

Le niveau situationnel comporte plusieurs paramètres : contexte social des activités communicatives (science, commerce, relations internationales, etc.), lieu, temps, nombre, rôle des locuteurs et relations entre eux.

Le niveau du contenu sémantique renvoie au sujet d'un texte, aux différentes perspectives (théorique, didactique, appliquée, de divulgation, etc.), aux parties textuelles (libres ou standardisées) et au développement thématique (séquences descriptives, narratives, expositives, argumentatives ou directives).

Le niveau formel-grammatical comporte les maximes rhétoriques, les formes linguistiques et non-linguistiques, les aspects grammaticaux, les ressources syntaxiques et lexicales.

Enfin, quand on parle d'un ensemble de textes spécialisés, on parle de *corpus*. Insistons sur le fait qu'un *corpus* en linguistique n'est pas constitué de textes rassemblés au hasard, mais bien au contraire de textes sélectionnés selon des critères bien définis, qui sont propres aux buts recherchés (cf. Meyer et Mackintosh 1996 ; Pearson 1998 ; Bowker et Pearson 2002).

5. VARIATION CONTEXTUELLE : TYPES DE CONTEXTES DE SPÉCIALITÉ

À la lumière des expériences actuelles en linguistique de *corpus*, y compris les nôtres depuis des années, une approche des terminologies et des langues de spécialité bien fondée doit, à nos yeux, commencer tout d'abord par la prise en considération de la variété des discours de spécialité et variations correspondantes. Les langues spécialisées ne sont pas des instruments monolithiques de communication, mais elles sont, par contre, bel et bien faites d'une pluralité de discours. Un deuxième principe de base pour une approche linguistique des terminologies et des langues de spécialité concerne la prise en compte, dans le cadre d'un certain discours de spécialité, de la variété de types ou genres textuels conditionnant souvent les ressources linguistiques choisies par l'énonciateur en fonction du destinataire, y compris les unités lexicales et phraséologiques, entre autres. Enfin, un troisième principe de base fondamental dans toute approche linguistique des terminologies et des langues spécialisées concerne la prise en compte de la variation contextuelle. Une notion centrale de la pragmatique est la notion de contexte ou plutôt l'explicitation des différents types de contexte. La notion de contexte en terminologie n'est pas toujours claire. Elle renvoie souvent tout simplement au contexte linguistique, phrastique ou microcontexte, notamment dans le cadre de la linguistique de *corpus* informatisés et de leur traitement automatique. Pourtant, la prise en compte des différents types de contexte est l'un des acquis les plus importants des études pragmatiques (cf. Armengaud 1990 : 62). Nous distinguons quatre grands types de contextes, que nous considérons comme étant une bonne base de travail pour la « linguistique de spécialité » et qui peuvent être présentés comme suit :

<i>II. Contextes de spécialité</i>	
Contexte circonstanciel, factuel, existentiel ou référentiel	Identité des locuteurs, environnement physique, lieu et temps où les propos sont tenus (contexte qui contient les individus dans le monde réel)
Contexte situationnel ou paradigmatique	Contexte qui détermine des rôles illocutionnaires plus ou moins institutionnalisés (p. ex. : une communication dans un congrès)
Contexte interactionnel	Enchaînement des actes de langage dans une séquence interdiscursive. Les interlocuteurs tiennent des rôles proprement pragmatiques : <i>proposer, objecter, rétracter</i> .
Contexte présuppositionnel	Contexte constitué par tout ce qui est également présumé par les interlocuteurs, leurs présuppositions ou leurs croyances, ainsi que leurs attentes et leurs intentions.

Tableau 3 : Types de contextes de spécialité

La terminologie se voit dans le besoin d'intégrer ces différents types de contextes — et non seulement le contexte linguistique ou microcontexte — à ses modèles d'analyse et de description.

6. TERMINOLOGIE ET VARIATION : LEXIQUE, SYNTAXE, SÉMANTIQUE ET PRAGMATIQUE (LSSP)

Le modèle théorique que nous proposons pour le traitement des terminologies et des langues de spécialité est un modèle foncièrement linguistique, bien qu'il prenne aussi en compte l'extralinguistique. Nous préférons la notion d'unité lexicale spécialisée plutôt que celle de terme, dans la mesure où la première limite les effets de *corpus* (candidats termes) et les effets d'ontologies (*cf.* Lerat 2002), qui conduisent à la marginalisation de ce qui n'est pas nom ou groupe nominal. La notion d'unité lexicale spécialisée renvoie, tout d'abord, à des sens et à des emplois spécialisés.

La recherche en terminologie et LSP ne peut pas se limiter au seul niveau lexical, et encore moins au seul niveau de la classe grammaticale nominale. Pour saisir le fonctionnement (sens et emplois) des unités lexicales spécialisées, il faut les analyser et les décrire dans leur fonctionnement morphosyntaxique et syntaxique. Sur le plan de la phrase, analysée en extension, on arrive à saisir la variation syntagmatique et la variation paradigmatique, pour arriver à cerner la variation conceptuelle et dénominate. On est là sur le plan de la variabilité inhérente à toutes les langues, celle qui se manifeste au niveau des variations lexicales, syntaxiques et sémantiques, intralinguistiques et interlinguistiques. On reste au niveau des emplois.

Cependant, les phrases « spécialisées » ne se forment pas dans le vide, mais elles surgissent et composent les textes de spécialité, ceux-ci formant des discours de spécialité. Lorsqu'on dépasse le plan phrastique, on entre dans le plan textuel. Or, comme le prouvent les travaux en linguistique textuelle, dans toute langue naturelle, trois axes extralinguistiques déterminent la production textuelle et la variation linguistique : l'axe socioculturel ou sociolectal ; l'axe géographique ou topolectal ; l'axe chronologique ou chronolectal. D'un point de vue linguistique, ces axes renvoient aux variations linguistiques liées aux différents types de locuteurs, aux variations linguistiques dépendant de l'espace ou des régions du globe, aux variations linguistiques liées aux périodes successives dans le temps. On est sur le plan des usages.

Les langues en tant que vecteurs de connaissances spécialisées, autrement dit les langues de spécialité, n'échappent pas à ces variations qui, d'un point de vue linguistique, se manifestent sur le plan textuel. Cependant, la relation étroite et intrinsèque entre l'unité lexicale spécialisée et son contexte supraphrastique, textuel et discursif, demeure ignorée, dans la mesure où jusqu'à présent elle n'a jamais été véritablement intégrée dans un modèle linguistique global d'analyse et de description des langues de spécialité.

C'est dans le but de combler les lacunes des approches partielles des LSP, que nous proposons un modèle d'analyse large, intégrant les plans d'analyse linguistique majeurs : lexical, syntaxique, sémantique et pragmatique. Théoriquement, une approche globale ne peut que comprendre à la fois sens, emplois et usages. Le plan pragmatique conduit à la prise en compte de la variation discursive et textuelle conditionnant le sens et l'emploi des unités lexicales spécialisées.

Du point de vue syntaxique et sémantique, les langues spécialisées ont tout à gagner à être décrites avec des outils méthodologiques qui ont fait leur preuve en linguistique générale. Mais la lexicographie générale et la lexicographie spécialisée ont tout à gagner à intégrer les usages textuels et discursifs dans leurs modèles de description du lexique spécialisé. Ainsi, une focalisation souhaitable dans les études spécialisées sera l'approche du lexique (noms, adjectifs, verbes et adverbes) via les variations textuelles et discursives.

7. EXEMPLIFICATION DU MODÈLE LLSP : VERBES ET VARIATIONS TEXTUELLES ET DISCURSIVES

Pour saisir le sens des mots dans les langues de spécialité, il est nécessaire, tout d'abord, de rendre compte de leurs emplois, au sens de G. Gross (1998 : 104-109) : « un schéma de phrase prédicat-arguments, une traduction, une réalisation morphologique (pour un prédicat, un verbe, un nom ou un adjectif), une actualisation (avec des degrés de figement) et un domaine, en particulier » (Lerat 2002 : 201). Tels sont les principes du modèle du Lexique-Grammaire du LLI et, en partie, ceux du modèle de la LEC d'Igor Mel'čuk, appliqués aux langues de spécialité. Partons donc de ces travaux et voyons leurs résultats et leurs limites.

Prenons des exemples du droit commercial, domaine qui relève à la fois du droit et du commerce. Les recherches dans ce domaine se trouvent à l'origine de l'élaboration du dictionnaire électronique bilingue des contrats du commerce international – français-portugais et portugais-français (cf. Desmet 2004). Nous nous limiterons ici à présenter des exemples en français.

Dans le cadre général des contrats du commerce international, nous trouvons le sous-domaine des contrats susceptibles d'être conclus, dans le cas d'installation d'une société à l'étranger. Ces contrats se subdivisent en plusieurs types, parmi lesquels les contrats de société, les contrats de bail, les contrats de marché de travaux, les contrats de coopération, les contrats de licence ou de brevet, etc.

Dans le cadre des contrats de bail, un verbe récurrent dans notre *corpus* est le verbe *louer*, un verbe polysémique de la langue générale (à ne pas confondre avec l'homonyme *louer – adresser des louanges*), employé dans une langue de spécialité. À première vue, il ne présente pas un intérêt particulier du point de vue de sa description dans un dictionnaire terminologique traditionnel. Dans un dictionnaire terminologique un peu plus « riche », autrement dit comportant des informations phraséologiques, il pourra apparaître dans le champ des collocations, sous l'entrée nominale « locaux » (entrée nominale, elle aussi à première vue peu spécialisée).

Dans un modèle visant le traitement de toutes les unités lexicales de spécialité (noms, verbes, adjectifs et adverbes), il sera pris comme une unité à part entière et pas seulement comme une « unité collocative ».

Si on applique la description postulée par le modèle du Lexique-Grammaire, nous aurons les données suivantes pour le terme français *louer*, dans le cadre des contrats de bail :

N0	Verbes	Prép1	N1	Prép2	N2	Synonymes	Dérivés
Humcol : société	louer		Inc : locaux	A	Humcol : société	donner en location	location
Humcol : société	louer		Inc : locaux	A	Humcol : société	prendre en location	location

Tableau 4 : Analyse du verbe « louer » dans le domaine du commerce international, selon le modèle du Lexique-Grammaire du LLI

(Légende : *Humcol* : humain collectif ; *inc* : inanimé concret)

Dans le cadre du commerce international, *une société loue des locaux à une autre société* (arguments typiques du verbe *louer* dans le cadre du domaine). Une telle description permet de mettre en évidence les arguments typiques dans le cadre général du domaine et de désambiguïser la forme verbale *louer*.

Mais dans les contrats du commerce international, une société qui donne des locaux en location est un *bailleur*, donner des locaux en location se dit *donner à bail* et la société qui prend des locaux à bail est un *locataire*. Ainsi, nous pouvons conclure que ce modèle mérite des affinages sur le plan de la description des arguments, en fonction des sous-domaines, s'il vise l'interprétation et la production de phrases.

N0	Verbes	Prép1	N1	Prép2	N2
Hum : bailleur	loue		Inc : locaux	à	Hum : locataire
Hum locataire	loue		Inc : locaux	à	bailleur

Tableau 5 : Analyse d'un deuxième emploi du verbe « louer » dans le domaine du commerce international, selon le modèle du Lexique-Grammaire du LLI

Nous pouvons conclure qu'en français il existe deux « emplois » de *louer* dans le cadre du domaine du commerce international, dans la mesure où ils font appel à des arguments de nature différente, à des arguments moins spécialisés dans le cas du premier emploi et à des arguments plus spécialisés dans le cas du second emploi.

Comme il est possible de le constater, ces analyses permettent de mettre à plat les constructions phrastiques typiques du domaine, les emplois possibles d'un verbe dans le cadre du domaine. Elles permettent donc de produire des phrases correctes dans le cadre du domaine. Toutefois, elles ne permettent pas de savoir dans quel contexte supraphrastique on doit utiliser ces « emplois ». À l'œil nu, on peut facilement identifier un emploi plus spécialisé et un emploi moins spécialisé, mais cela reste insuffisant. Il manque donc des indications d'usage dans les textes et dans les discours. Pour les obtenir, il faut remonter à un niveau supérieur à celui de la phrase, le niveau textuel. On peut donc se demander dans quels types de textes apparaît la forme verbale française *louer* accompagnée de ses arguments les plus spécialisés. Le *corpus* textuel nous répond à ces questions : le premier emploi de *louer* apparaît essentiellement dans les textes informatifs de la presse écrite générale et semi-spécialisée et dans les manuels spécialisés, alors que son deuxième emploi apparaît systématiquement dans les textes directifs, contrats et textes législatifs. En remontant encore plus haut, nous pouvons dire que *louer*₁ appartient au discours de divulgation, de semi-divulgation et technique didactique, alors que *louer*₂ apparaît dans le discours technique officiel (les contrats proprement dits) et le discours législatif (codes du commerce international). La variation textuelle et discursive est au cœur même de notre proposition de modèle théorique. Nous appellerons notre modèle « Lexique–Syntaxe–Sémantique–Pragmatique » (LSSP). Du point de vue de sa représentation, il est possible d'envisager l'ajout de ces informations de la manière suivante :

a) 1^{er} emploi de *louer* et usages

N0	Verbes	Prép.1	N1	Prép.2	N2	Synonymes	Dérivés	Textes	Discours
Humcol : société	Louer		Inc : locaux	À	Humcol : société	donner en location	location	Inf.	Vul. ; s-Vul ; Did.
Humcol : société	Louer		Inc : locaux	À	Humcol : société	prendre en location	location	Inf.	Vul. ; s-Vul ; Did.

Tableau 6 : Premier emploi du verbe « louer » et usages, dans le domaine du commerce international, selon le modèle « Lexique-Syntaxe-Sémantique-Pragmatique » (LSSP)

(Légende : *Inf.* : texte informatif ; *Vul.* : discours de vulgarisation ; *s-vul.* : discours de semi-vulgarisation ; *Did.* : discours technique didactique)

b) 2^e emploi de *louer* et usages

N0	Verbes	Prép.1	N1	Prép.2	N2	Textes	Discours
Hum : bailleur	loue		Inc : locaux	à	Hum : locataire	Dir.	Spé Spé offic
Hum locataire	loue		Inc : locaux	à	bailleur	Dir.	Spé Spé offic

Tableau 7 : Deuxième emploi du verbe « louer » et usages, dans le domaine du commerce international, selon le modèle « Lexique-Syntaxe-Sémantique-Pragmatique » (LSSP)

(Légende : *Dir.* : texte directif ; *Spé* : discours spécialisé ; *Spé offic* : discours spécialisé officiel)

8. CONCLUSION

Les catégorisations des textes et des discours peuvent éventuellement être encore plus fines, ainsi que leur codage pour une représentation formelle. En tout cas, les catégories

que nous avons identifiées nous semblent déjà être une information indispensable pour cadrer les emplois dans les usages, en intégrant ainsi des données pragmatiques essentielles pour la description des langues de spécialité.

Seules les indications de « valence pragmatique », autrement dit les usages dans les textes et dans les discours, peuvent départager les emplois.

Domaines, discours et textes doivent être combinés pour déterminer les usages. Les conséquences sont de taille, et pour la lexicographie générale et pour la terminographie.

Notre proposition de modèle Lexique–Syntaxe–Sémantique–Pragmatique–(LSSP) pour le traitement des verbes spécialisés peut être appliquée à toute unité lexicale de spécialité appartenant à n'importe quel domaine de spécialité.

BIBLIOGRAPHIE

- Armengaud (F.), 1990 : *La Pragmatique* (1985), 2^e éd., Paris, Presses universitaires de France, coll. « Que sais-je ? », n° 2230.
- Bowker (L.) et Pearson (J.), 2002 : *Working with Specialized Language, A Practical Guide to Using Corpora*, Routledge.
- Desmet (I.), 1998 : « Caractéristiques sémantiques, syntaxiques et discursives des vocabulaires spécialisés. Quelques données théoriques et pratiques pour la lexicographie spécialisée », dans *Actes du 2^e colloque de linguistique appliquée : « Les linguistiques appliquées et les sciences du langage »*, Strasbourg, COFDELA Publications, pp. 292-305.
- Desmet (I.), 2004 : « Terminographie d'apprentissage et apprentissage de la terminographie : le dictionnaire électronique bilingue des contrats du commerce international (portugais-français et français-portugais) », dans *Études de linguistique appliquée*, n° 135, Paris, Didier Érudition - Klincksieck, pp. 285-298.
- Gläser (R.), 1993 : « A Multi-level Model for a typology of LSP Genres », dans *Fachsprache, International Journal of LSP*, 1-2, pp. 18-26.
- Gross (G.), 1998 : « Pour une véritable fonction “synonymie” dans un traitement de texte », dans *Langages*, 131, pp. 103-114.
- Heinemann (W.) et Viehweger (D.), 1991 : *Textlinguistik : eine Einführung*, Tübingen, Niemeyer.
- Heinemann (W.), 2000 : « Textsorten. Zur Diskussion um Basisklassen des Kommunizierens. Rückschau und Ausblick », En Adamzik, K. (Od.), dans *Textsorten*, Tübingen.
- Hoffmann (L.), 1976 : *Kommunikationsmittel Fachsprache. Eine Einführung*, Berlin, Akademie-Verlag.
- Kocourek (R.), 1991 : *La langue française de la technique et de la science*, Paris/Wiesbaden, Oskar Brandstetter.
- Laurian, (A.-M.), 1983 : « Typologie des discours scientifiques : deux approches », dans *Études de linguistique appliquée*, n° 52, Paris, Didier Érudition, pp. 8-20.
- Lerat (P.), 1995 : *Les Langues spécialisées*, Paris, PUF.
- Lerat (P.), 2002 : « Qu'est-ce qu'un verbe spécialisé ? Le cas du droit », dans *Cahiers de Lexicologie*, n° 80, Paris, Honoré Champion, pp. 201-211.
- Mel'čuk (I.) et al ; 1995 : *Introduction à la lexicologie explicative et combinatoire*, Éditions Duculot/AUPELF-UREF.
- Pearson (J.), 1998 : *Terms in Context*, Amsterdam, John Benjamins Publishing Company.
- Meyer (I.) et Mackintosh (K.), 1996 : « The Corpus from a Terminographer's Viewpoint », dans *International Journal of Corpus Linguistics* 1 (2), pp. 257-285.

Rey (A.), 1979 : *La Terminologie, noms et notions*, Paris, Presses Universitaires de France, Que Sais-je ?, n° 1780.

Weise (G.), 1993 : « Criteria for the classification on ESP texts », dans *Fachsprache. International Journal of LSP*, 1 / 2, pp. 26-31.

LES ADVERBES COMPOSÉS DANS LE DOMAINE DU TRAVAIL

Dolors Català,
departament de filologia francesa i romànica,
Universitat Autònoma de Barcelona,
08193 Bellaterra
Barcelona, España,
dolors.catala@uab.es
Jorge Baptista,
departamento de letras clássicas e modernas,
Universidade do Algarve,
Campus de Gambelas,
8005-139 Faro, Portugal
jbaptis@ualg.pt

Résumé : *La description des langues de spécialité se limite habituellement à la terminologie nominale et néglige souvent l'importance des adverbes, qui sont aussi des prédicats structurants du vocabulaire. Notre travail s'inscrit dans le cadre méthodologique du lexique-grammaire et de la notion de classes d'objets. Dans cette communication, nous présenterons le cadre général des problèmes posés par la description des adverbes composés dans les langues de spécialité et notamment ceux qui appartiennent au domaine du travail dans des langues voisines espagnol/français/portugais ainsi que quelques ressources linguistiques informatiques que nous avons développées.*

Mots-clés : *adverbe composé, dictionnaire électronique, traduction automatique, contexte, travail, profession*

1. INTRODUCTION

Le lexique et la grammaire des langues de spécialité ayant pour objectif la traduction automatique suscitent un intérêt particulier chez les linguistes, intérêt facilement intégrable à une attention pour le social. En effet, la recherche d'informations dans des sources documentaires, notamment sur le web, est de plus en plus fréquente et exige des outils linguistiques appropriés, et la technologie employée actuellement n'inclut pas de modules assez performants en matière de récupération et de filtrage de l'information.

La description des langues de spécialité se limite habituellement à la terminologie nominale. Le terminologue-terminographe s'attache principalement à délimiter et à définir des concepts représentés par des termes. Ses fiches terminologiques donnent priorité aux données de type encyclopédique comme le domaine et la définition, de type signalétique (les sources documentaires) et de type traductologique (équivalents en une

ou plusieurs langues cibles). Mais le terme, pour sa part, demeure une réalité linguistique dont le fonctionnement syntactico-sémantique est rarement abordé ou même simplement ignoré. Les verbes et les adjectifs, quand ils sont pris en compte, se trouvent confinés à des champs ressentis comme secondaires à l'intérieur de la fiche. On néglige souvent l'importance des adverbes, qui sont aussi des prédicats structurants du lexique. Nous nous sommes penchés sur les adverbes composés et nous avons élaboré un dictionnaire électronique espagnol/français/portugais en vue de la traduction automatique.

Dans le cadre de cet article, nous nous sommes particulièrement intéressés aux adverbes composés utilisés dans le domaine du travail. Dans ce domaine, nous trouvons des adverbes concernant les conditions contractuelles qui apparaissent dans des phrases du type :

ES : [*trabajar*] *a tiempo (parcial + completo)*

FR : [travailler] à (temps partiel + plein temps)

PT : [*trabalhar*] *a tempo (parcial + integral)*

ES : [*ser pagado*] *a comisión*

FR : [être payé] à la commission

PT : [*ser pago*] *à comissão*

On trouve aussi des adverbes qui modifient des prédicats spécifiques à certaines activités professionnelles comme :

– Médecine :

ES : [*operar*] *con anestesia epidural*

FR : [opérer] sous anesthésie épidurale

PT : [*operar*] *con anestesia epidural*

– Musique :

ES : [*cantar*] *a cappella*

FR : [chanter] a cappella

PT : [*cantar*] *a cappella*

– Économie :

ES : [*comprar*] *< mercancía > al por (mayor + menor)*

FR : [acheter] < marchandise > (en gros + au détail)

PT : [*comprar*] *< mercadoria > a (grosso + granel + retalho)*

– Culinaire :

ES : [*cocinar*] *< alimento > a fuego (lento + moderado + fuerte)*

FR : [cuisiner] < aliment > (à petit feu + à feu modéré + à feu vif)

PT : [cozinhar] <alimento> em lume (brando + moderado + forte)

Il s'agit, pour la plupart, d'adverbes composés qui posent des problèmes particuliers pour l'établissement des correspondances entre des langues voisines mais aussi pour la définition des formats de dictionnaires électroniques en tant qu'outils d'aide à la traduction.

Dans cette communication, en premier lieu, nous présenterons brièvement le cadre théorique qui sous-tend notre recherche. En deuxième lieu, nous nous pencherons sur les adverbes dans les langues de spécialité, puis nous proposerons quelques outils informatiques que nous avons développés.

2. CADRE THÉORIQUE

2.1. Lexique-grammaire des phrases élémentaires

Pour faire une description de la langue de spécialité, nous avons adopté la démarche du lexique-grammaire élaboré par Maurice Gross (1981, 1990a) et développé par Gaston Gross (1994d, 1995b, 1996b) notamment pour les questions relatives à la sémantique. Dans ce cadre basé sur la théorie transformationnelle de Z. S. Harris (1976), l'unité minimale de sens n'est pas le mot mais la phrase simple, qui s'analyse en termes de prédicat et d'arguments selon le schéma d'arguments suivant :

prédicat (argument 1, argument 2, argument 3)

Le prédicat est l'élément central de la phrase et les arguments sont les éléments indispensables pour que la suite soit grammaticale.

2.2. Lexique-grammaire de l'adverbe

En ce qui concerne les adverbes, nous prenons appui sur les travaux de Maurice Gross (1986). Il applique la notion d'adverbe à des structures syntaxiquement différentes de la terminologie traditionnelle comme les adverbes proprement dits, les propositions subordonnées circonstancielles, les compléments circonstanciels et d'autres formes encore, telles que les exclamations et les conjonctions ayant des propriétés adverbiales. Il présente une classification des formes des adverbes et non de leur fonction. Il distingue les adverbes figés des adverbes libres et constate un continuum entre ces deux types de suites. Les adverbes composés ont souvent été considérés comme des exceptions et ont été habituellement écartés des discussions théoriques. Ils constituent pourtant une partie importante du lexique. Maurice Gross et son équipe du laboratoire d'automatique documentaire et linguistique (LADL, Université Paris VII) en ont répertorié et classé, d'après leur forme et leurs propriétés syntaxiques, environ 7 000 pour le français (Gross 1986). Suivant les mêmes critères méthodologiques, 6 000 constructions adverbiales de l'espagnol (Català 2003) et environ 2 000 du portugais (Palma, travail en cours) ont été inventoriées.

2.3. Les classes d'objets

Nous nous appuyons aussi sur la notion de « classes d'objets » introduite par Gaston Gross (1992) comme outil permettant de structurer le lexique d'un point de vue syntactico-sémantique. Le principe de base consiste à classer les prédicats et les arguments moyennant des ensembles et des sous-ensembles sémantiquement homogènes constitués à partir de propriétés lexico-syntaxiques communes. En effet, les traits sémantiques traditionnellement utilisés sont insuffisants pour délimiter le sens d'une phrase, ainsi pour rendre compte de l'emploi du verbe *garer* il ne suffit pas de dire qu'il implique un sujet *humain* < hum > et un objet *inanimé concret* < inc > :

< hum > garer < inc >

car cela donnerait lieu à des phrases du genre :

* l'étudiant gare ses livres

* le boucher gare sa viande

Il est nécessaire de dire que le verbe *garer* implique un nom de *moyen de transport* < inc : mtr > en position de complément :

< hum > garer < inc : mtr >

Les propriétés distributionnelles d'un verbe comme *garer* constituent un instrument important de la description linguistique, car elles permettent de le considérer comme un opérateur approprié à une classe sémantiquement homogène de noms (les noms de moyens de transports, dans ce cas). Une fois cette classe de noms constituée, elle peut aussi être utilisée pour décrire de façon plus précise la distribution d'autres prédicats.

D'autre part, la structuration en classes d'objets peut être opératoire pour préciser les propriétés syntaxiques ou pour la traduction dans la mesure où elles permettent de monosémiser les prédicats, comme nous pouvons l'observer dans l'exemple suivant. Considérons l'adjectif *léger* du français, tel qu'il est employé dans les constructions suivantes :

FR :	blessure/blessé	< léger >
ES :	<i>herida/herido</i>	< leve > /*< ligero >
PT :	<i>ferida/ferimento/ferido</i>	< leve > /< ligeiro >
FR :	faute	léger
ES :	<i>falta</i>	leve/*ligera
PT :	<i>falta</i>	leve/ligeira

Dans ces constructions, *léger* s'oppose *grosso modo* à *grave*. Pourtant, il semble y avoir une subtile différence entre les deux emplois de *léger*, du fait qu'avec *blessure* le concept de « superficiel » y est aussi attaché. Cette différence peut ressortir de façon

plus nette dans la recherche d'une traduction adéquate. Les solutions présentées par chaque langue peuvent être distinctes.

Ainsi, bien que l'espagnol ait l'adjectif *ligero*, historiquement associé à l'adjectif français, il l'associe rarement à *herida* et pratiquement jamais à *falta*, préférant donc l'adjectif *leve*. Par contre, en portugais, les deux adjectifs *leve* et *ligeiro* sont permis, mais on remarque une tendance à utiliser plutôt *ligeiro* que *leve* avec *ferida*. Inversement, avec le nom *falta*, *leve* est plus usuel que *ligeira*.

D'autre part, certaines combinaisons doivent être codées dans des dictionnaires, si on ne veut pas aboutir à des traductions fautives. Par exemple, *voiture légère* constitue sans doute une combinaison figée qui ne peut pas se traduire avec l'adjectif *leve*, ni en ES ni en PT.

2.4. Les classes d'objets et l'adverbe

L'adverbe (simple ou composé), parmi ses diverses fonctions, exerce le plus souvent une prédication (modification) sur un verbe ou un adjectif, ou encore sur une phrase complète. Dans la perspective harrissienne, il constitue, de ce fait, un opérateur d'ordre supérieur.

Or, les contraintes distributionnelles que l'adverbe impose aux prédicats qui sont ses arguments peuvent servir aussi à caractériser des classes syntactico-sémantiques de prédicats. Ainsi, de la même façon que le verbe *garer*, comme nous avons vu plus haut, peut servir à délimiter une classe d'objets – c'est-à-dire, une classe sémantiquement homogène de noms qui fonctionnent comme des arguments de cet opérateur approprié –, un adverbe peut aussi fonctionner comme opérateur approprié d'une classe sémantiquement homogène de prédicats (ses arguments). C'est le cas, par exemple, des adverbes :

FR : à pas (de tortue + de géant + de course + de loup + vifs + lents + légers)

ES : *a paso* (de tortuga + de carga + de buey + largo + ligero) + *a pasos agigantados*

PT : *a passo* (de tartaruga + de gigante + de corrida + de caracol + rápido + acelerado + incerto)

qui ne modifient que des verbes (ou des prédicats) de mouvement dynamiques < *vmvt* : *dyn* >, indépendamment de leurs emplois littéraux ou figurés, comme l'illustrent les exemples suivants¹ :

eficios va a reportar a la persona obesa. Caminar a paso ligero, como mínimo una hora diaria, facilitita

¹ Les concordances de l'espagnol sont extraites de la page web de la Real Academia Española (<http://www.rae.es>). Les concordances du portugais sont extraites du corpus *CETEMPúblico*, qui est rendu accessible par Linguateca (www.linguateca.pt).

<i>a en las despedidas, y dando media vuelta regresó cándalo. Poco después, la</i>	a paso largo <i>en busca del es-</i>
<i>más que nada un ejercicio de paciencia. Avanzaban ponía el sol debían det</i>	a paso de tortuga y <i>apenas se</i>
<i>ver. Ahí cruza raudo el caminante. Miento, avanza queante, parsimonioso, con</i>	a paso de buey. <i>Lento, ren-</i>
<i>ia a la puerta principal de la residencia, venían acero oscuro, de tama</i>	a paso de carga <i>dos dogos color</i>
<i>or Internet, se llega a creer que el mundo camina completo desorden. Como</i>	a pasos agigantados <i>al más</i>
<i>campo do Varzim e continua a sua marcha ão de Honra.</i>	a passo acelerado <i>para a Divis-</i>
<i>4 quilómetros e nosso caminhante marcha dos Olivais para não perd</i>	a passo apressado <i>para a Doca</i>
<i>e a liberalização dos serviços continua</i>	a passo de caracol. <i>Já</i>
<i>am-se a-dois e lá vão, para cima, do primeiro-ministro.</i>	a passo de corrida <i>para a porta</i>
<i>s encostadas ao bar. Atravessou a sala pares absortos num vel</i>	a passo incerto, <i>acotovelando os</i>
<i>a 700 contos, enquanto a IBM prossegue to da sua linha PowerPC,</i>	a passo rápido <i>o desenvolvimen-</i>
<i>emplo, que o abuso da heroína « progride conquista terreno, o núm</i>	a passos de gigante, <i>o tráfico</i>
<i>o vento./Vi como a noite veio,/Veio pesadas cicutas/Senti m</i>	a passos largos <i>como a cor das</i>
<i>Tudo leva assim a crer que se caminha pendência das manifestações</i>	a passos rápidos <i>para a inde-</i>

Pourtant, bon nombre d'adverbes composés présentent le plus souvent une distribution très contrainte, voire unique, dans le sens qu'ils se combinent de façon exclusive avec un seul prédicat ou une toute petite famille de prédicats, sémantiquement très voisins. C'est le cas des adverbes suivants porteurs d'une valeur augmentative :

FR : [rire] à gorge déployée

ES : [reir] a carcajada tendida

PT : [rir] às bandeiras despregadas

mais qui ne se combinent qu'avec le verbe *rire*.

De même, nous trouvons aussi des adverbes ayant des distributions assez contraintes et appartenant à des domaines spécifiques. Par exemple, dans le domaine culinaire, on pourrait citer la petite famille des adverbes qui spécifient l'intensité des processus de cuisson :

FR : [cuisiner] < aliment > à feu (doux + modéré + vif) + à petit feu

ES : [cocinar] < alimento > a fuego (flojo + lento + moderado + fuerte + suave + vivo)

PT : [cozinhar] < alimento > em (fogo + lume) (brando + esperto + vivo + forte + fraco + moderado)

mantequilla y saltear la cebolla en esta (sofreír) a fuego lento. *Agregar la harina, moviendo constantem*

combine la miel, agua y jugo de limón y deje hervir a fuego lento. *Agregue la canela y trozos de guayaba.*

o se reparta uniformemente. Déjela cocer 1 minuto a fuego fuerte, *sin remover,* *retírela y esparza el qu*

cacerola y rehogar en ella la cebolla y el apio, a fuego suave, *sin permitir* *que tomen demasiado color*

ollas, salpimentar los filetes de pescado y dorar a fuego vivo *con un poco de* *aceite. — Pelar las cebol*

os legumes, e o princípio o da cozedura em lume brando. *O que deu* *os vários cozidos de Portuga*

untada de azeite, cozinhe as costeletas em lume forte, *durante três* *minutos de cada lado. Junt*

On remarquera que certains de ces adverbes composés appartenant à un domaine spécifique peuvent, par plusieurs processus métaphoriques, s'autonomiser de leurs emplois originels. C'est le cas, par exemple, de l'adverbe PT : *em lume brando*, qui est devenu d'usage courant :

a polémica prosseguiu em lume brando. *Na Mealhada, foi criada a* *Associação*

Para alimentar em lume brando *uma questão de que pretende* *retirar benefícios*

Trinta anos de amadurecimento em lume brando *da imagem de Blueberry,*

On peut aussi trouver en espagnol les deux expressions *a fuego vivo* et *a fuego lento* dans un sens métaphorique :

No quise descartar la idea sin antes hervirla a fuego vivo, *ni mencionársela a* *mis padres*

Todo lo contrario : un hombre cordial, forjado a fuego vivo, que había subido por la escalera del bu

de CiU, Jordi Xuclà, abogó por una reflexión "a fuego lento" y tomando en cuenta la opinión de los

Dans ce nouvel emploi, qui constitue, donc, une nouvelle entrée lexico-syntaxique, les oppositions de l'emploi culinaire ne marchent plus :

* *A polémica prosseguiu em lume forte/vivo/esperto*

(La polémique s'est poursuivie à feu vif)

Nous reviendrons sur ces transferts métaphoriques dans la section suivante.

3. L'ADVERBE COMPOSÉ DANS LES LANGUES DE SPÉCIALITÉ

Les notions que nous venons de présenter ont été élaborées et appliquées dans le cadre de la langue générale. Ce même modèle peut nous servir à décrire les langues de spécialité. L'introduction des classes d'objets dans l'analyse des langues de spécialité n'en est encore qu'à ses débuts, c'est pourquoi il nous semble nécessaire d'approfondir les recherches dans ce domaine. Nous considérons en suivant Lerat (1995) que la spécificité des langues de spécialité réside plutôt dans les emplois spécialisés et les usages que l'on fait de la langue générale dans des contextes particuliers.

Dans cette optique, si nous examinons par exemple un *corpus* de textes médicaux, il est possible de distinguer plusieurs types d'unités adverbiales :

1. Des adverbes strictement spécialisés : *in vitro, sous anesthésie générale* :

ES : *El cirujano ha extirpado el tumor con anestesia general*

FR : Le chirurgien a extirpé la tumeur sous anesthésie générale

PT : *O cirurgião extraiu o tumor sob anestesia geral*

ES : *Los médicos fecundan los óvulos in vitro*

FR : Les médecins fécondent les ovules in vitro

PT : *Os médicos procedem à fecundação dos óvulos in vitro*

2. Des emplois spécialisés d'adverbes généraux : *à froid/à chaud* :

ES : *El cirujano opera el paciente en (frío + en caliente)*

FR : Le chirurgien opère le patient à (froid + chaud)

PT : *O cirurgião operou o paciente a (frio + quente)*

3. Des adverbes généraux non spécifiques du domaine, comme *en profondeur*, *sous surveillance (médicale)* :

ES : *El tumor se extiende en profundidad*

FR : La tumeur s'étend en profondeur

PT : *O tumor estende-se em profundidade*

ES : *El paciente toma medicamentos bajo estricta vigilancia (médica)*

FR : Le patient prend des médicaments sous surveillance (médicale)

PT : *O paciente toma os medicamentos sob (estrita) vigilância (médica)*

Certains adverbes spécialisés du domaine médical apparaissent dans la langue générale et correspondent à des emplois métaphoriques. C'est le cas des adverbes :

ES : *en dosis homeopáticas*

FR : en doses homéopathiques

PT : *em doses homeopáticas*

comme nous pouvons l'observer dans les exemples suivants :

ES : [...] *su tesis, pura y dura, era que la enseñanza en catalán -aunque fuera en dosis homeopáticas- a los alumnos castellano parlantes había de ser absolutamente voluntaria* [...]

<http://www.arrakis.es/~corcus/losantos/debate2300/18b.htm> (12.08.2005 ; 13:32)

(sa thèse, pure et dure, consistait en ce que l'enseignement en catalan – bien que ce soit en doses homéopathiques – à des élèves parlant le castillan devrait être absolument volontaire)

FR : [...] Avant d'en arriver là où il est aujourd'hui, le plan d'épargne en actions ne s'est entrouvert *qu'à doses homéopathiques*. [...]

<http://www.lesechos.fr/patrimoine/bourse/200056793.htm> (12.08.2005 ; 13:40)

PT : [...] *Velha e sempre moderna, de trânsito caótico Roma é uma cidade fascinante que se deixa descobrir em doses homeopáticas*. [...]

<http://www.netviagens.iol.pt/>(12.08.2005; 13:45)

(Vieille et toujours moderne, à circulation chaotique, Rome est une ville fascinante qui se laisse découvrir en doses homéopathiques)

Ce n'est pas un phénomène isolé car bon nombre d'adverbes qui appartenaient à l'origine à des domaines spécialisés, notamment à celui des professions, sont passés par la suite à la langue générale après un processus de généralisation. C'est le cas, par

exemple, de l'adverbe ES : *al garete* utilisé par les marins signifiant « sous l'effet des vents », qui s'est figé dans la langue générale dans l'expression *irse al garete*, c'est-à-dire « aller au diable » (Seco *et al.* 2004). De même, les expressions équivalentes en français FR : < aller >/< être > à *la dérive* et en portugais PT : < *ir* >/< *estar* > à *deriva* étaient à l'origine des termes maritimes, mais elles sont passées à la langue générale, signifiant en français « se laisser aller sans réagir » (*Petit Larousse Illustré* 1992) et en portugais « sans contrôle ».

Nous pensons que tous ces types d'adverbes doivent faire l'objet d'une description lexicographique systématique et détaillée, ce qui nous permettra de fournir des outils lexicographiques pour la traduction automatique espagnol/français/portugais.

4. EXEMPLE D'UNE GRAMMAIRE LOCALE DU TRAVAIL

En vue de décrire les adverbes figés d'une langue de spécialité, nous nous sommes centrés sur ceux qui apparaissent dans le domaine du travail dans des phrases du type :

ES : *El hotel ha contratado un nuevo cocinero a media jornada*

FR : L'hôtel a engagé un nouveau cuisinier à mi-temps

PT : *O hotel contratou un novo cozinheiro a meio tempo*

Il s'agit d'adverbes désignant des conditions contractuelles. On trouve plusieurs séries d'adverbes concernant, par exemple, des contrats (temporaires) dont la rémunération est centrée sur les résultats :

ES : *Este joven es pagado a (la pieza + destajo)*

FR : Ce jeune est payé à (la pièce + forfait)

PT : *Este rapaz é pago à (peça + tarefa)*

ou bien, sur le temps de travail :

ES : *Este joven es pagado (por horas + a (jornal + la semana))*

FR : Ce jeune est payé à (l'heure + la journée + la semaine)

PT : *Este rapaz é pago (à hora + ao dia + à semana)*

Naturellement, ces derniers peuvent être très idiomatiques et ne pas avoir d'équivalents dans les autres langues, car ils relèvent des normes juridiques particulières du pays :

ES : « *Es decir, que una cafetería no debería contratar por obra a un camarero para servir sólo café en la barra* »

<http://www.elmundo.es/sudinero/noticias/act-85-5.html>

FR : « Si mon travail vaut 110 pour cent du SMIC, on pourra me payer au SMIC »

<http://www.atoute.org/dcforum/DCForumID20/2254.html>

PT : *A empresa contratou um funcionário a recibos verdes*

À ces adverbess, il faudra aussi ajouter des expressions spécialisées décrivant la résiliation d'un contrat. Par exemple :

PT : *A empresa rescindiu amigavelmente o contrato com o funcionário*

FR : L'entreprise et le salarié ont résilié le contrat à l'amiable

ES : *La empresa ha rescindido amigablemente un contrato con el empleado*

PT : *A empresa despediu o funcionário (com + sem) justa causa*

ES : *La empresa ha despedido el empleado (con + sin) causa justificada*

FR : L'entreprise a renvoyé l'employé sans cause justifiée

Remarquons que bon nombre de ces unités s'appliquent souvent sur des noms prédicatifs accompagnés de leur verbe support qu'il faut prendre en compte (Gross 1981). Ainsi, pour l'expression PT : *rescindir amigavelmente < contrato >*, de l'exemple ci-dessus, nous avons aussi pris en considération la nominalisation associée *fazer uma rescisão de < contrato >*, où l'adverbe en -ment (*amigavelmente*), en tant que modifieur du nom prédicatif *rescisão* « résiliation », se transforme en l'adjectif (*amigável*) morphologiquement associé (Giry-Schneider 1978, 1987) :

PT : *A empresa fez uma rescisão amigável do contrato com o funcionário*

ES : *la empresa ha hecho una rescisión amistosa del contrato*

FR : l'entreprise a fait une résiliation amicale du contrat avec l'employé

Nous avons construit des grammaires locales regroupant ces ensembles d'expressions, sous forme d'automates à états finis (*directed acyclic graphs*). Pour la construction des graphes et pour l'exploitation des *corpus*, nous avons utilisé le système Intex (Silberztein 1993, 2000), avec les dictionnaires électroniques du français (Courtois, Silberztein 1990), de l'espagnol (Blanco 2000) et du portugais (Ranchhod *et al.* 1999). Ces automates ont été ensuite appliqués à des *corpus* de langue générale, ce qui nous a permis de vérifier l'ensemble des prédicats (et leurs dérivés) sur lesquels portent ces adverbess. À titre d'exemple, nous commentons les résultats obtenus à partir d'un automate construit pour le portugais (figure 1).

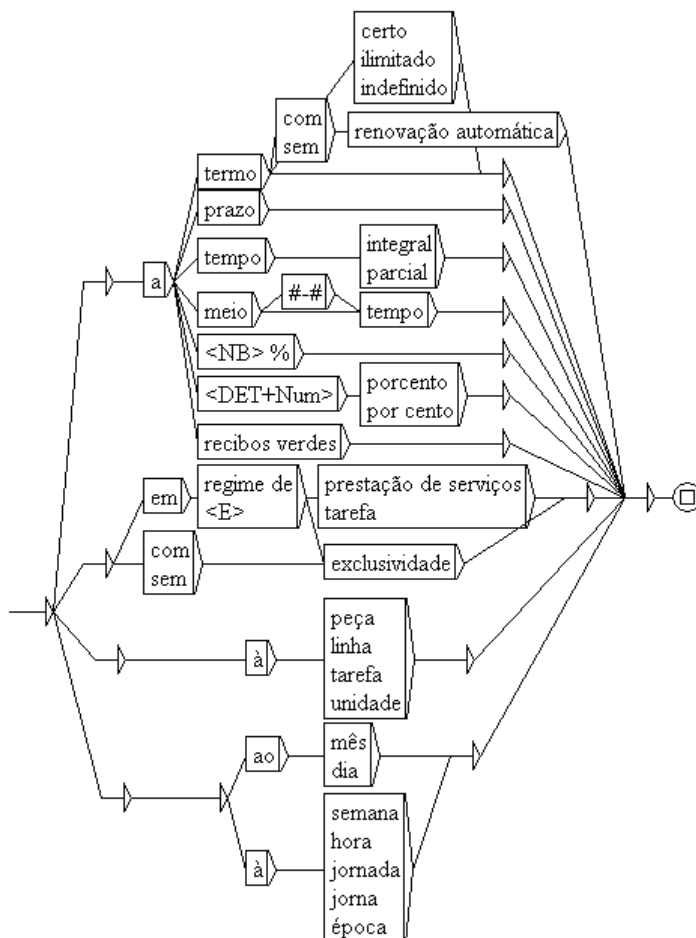


Figure 1 : Grammaire locale des adverbes composés désignant les < mode_de_contrats > du Portugais

L'application de ce graphe sur le corpus² nous a permis d'établir les regroupements sémantiques suivants qui sont en nombre remarquablement réduit (d'un ensemble de 1 147 « matches »). Entre autres, on trouve :

- *contrato ou contrato de trabalho, contratar, contratado* ; (contrat)
- *trabalho, trabalhar, trabalhador* ; (travail)
- *emprego, empregar, empregado* ; (emploi)
- *cargo, posto, lugar, posição, situação* ; (poste)

² CETEMPúblico, premier fragment (environ 9,6 millions de mots).

- *despedimento, despedir, despedido* ; (congédier)
- *rescisão, rescindir, rescindido* ; (résiliation)
- *pagamento, pagar, pago* ; (payer)
- *receber, recebido* ; (recevoir)
- *ganhar ; cobrar* ; (gagner, toucher)

Cependant, plusieurs problèmes d’ambiguïté formelle sont à l’origine des résultats fautifs. Ainsi, par exemple, l’adverbe PT : *a prazo* peut aussi faire partie d’une autre série d’adverbes qui appartiennent au domaine de la bourse et qui ne s’appliquent qu’à des < opérations_financières > ou des < produits_financiers > :

PT : *O Pedro fez um depósito a prazo* (Pedro a fait un dépôt)

Pour pallier ce problème, nous avons donc construit un graphe (Vprof) avec tous les prédicats (et leurs dérivés) associés à des adverbes du domaine du travail (ADVprof) et nous avons recherché des motifs en permettant l’insertion des mots quelconques entre les deux sous-graphes (boîtes grises), ce qui est représenté par le symbole < MOT > avec une boucle dans le graphe de la figure 2 :

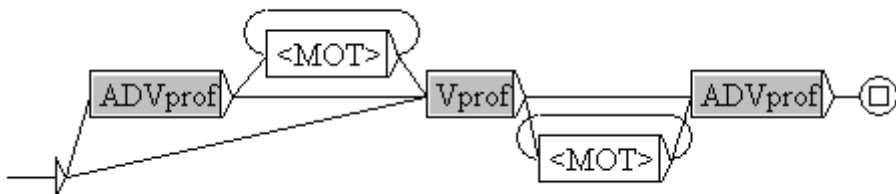


Figure 2 : Graphe pour la recherche des motifs des combinaisons de prédicats (*Vprof*) et les adverbes composés (*ADVprof*) du domaine du travail

Nous avons ainsi repéré 120 « matches », dont seulement 21 fautifs, ce qui correspond à un taux de succès de 82,5 %. En recherchant les mêmes motifs sans admettre des mots insérés entre les sous-graphes, nous obtenons 98 « matches », dont un seul fautif, c’est-à-dire, un taux de succès de 98,9 %. Ceci illustre bien l’utilité de ce type d’approche pour plusieurs applications dans la recherche documentaire (*information extraction*).

5. CONCLUSION

Pour décrire les adverbes composés de spécialité nous avons utilisé les mêmes outils que pour la langue générale en veillant à replacer ces unités dans leur contexte, la phrase. Nous avons souligné les difficultés de leur description en vue du traitement automatique et nous avons proposé des outils utiles pour la reconnaissance de ces unités en vue de la traduction automatique. Cette approche constitue un point de départ pour avancer dans la description des langues de spécialités en y incluant une composante non nominale.

Notations

hum = humain ;

inc = inanimé concret ;

mtr = moyen de transport

vmvt = verbe de mouvement

dyn = dynamique

< *mot* > représente toutes les flexions de *mot* (e. g. *mot* et *mots*)

BIBLIOGRAPHIE

- Blanco (X.), 2000 : « Les dictionnaires électroniques de l'espagnol, Delass et Delacs », *Lingvisticae Investigationes* 23, Amsterdam/ Philadelphia, John Benjamins, pp. 201-218.
- Català (D.), 2003 : *Adverbes composés. Approches contrastives*, thèse de doctorat, Barcelone, Univ. Autònoma Barcelona.
- Courtois (B.) Silberztein (M.), (1990) : « Dictionnaires électroniques du français », dans *Langue française* 87, Paris, Larousse.
- Giry-Schneider (J.), 1978 : *Les Nominalisations en français*, Genève, Droz.
- Giry-Schneider (J.), 1987 : *Les Prédicats nominaux en français. Les Phrases simples à verbe support*, Genève, Droz.
- Gross (G.), 1992 : « Forme d'un dictionnaire électronique », dans *L'Environnement traductionnel. La Station du traducteur de l'an 2001*, Sillery/Montréal, Presses de l'Université de Québec/AUPELF-UREF, pp. 255-272.
- Gross (G.), 1994 : « Un outil pour le FLE : les classes d'objets », dans *Actes du colloque FLE* de mai 1992, Presses Universitaires de Lille, pp. 169-186.
- Gross (G.), 1995 : « Une sémantique nouvelle pour la traduction automatique. Les classes d'objets », dans *La tribune des industries de la langue et de l'information électronique* 17-18-19, Paris.
- Gross (G.), 1996 : *Les Expressions figées en français*, Paris, Ophrys.
- Gross (M.), 1981 : « Les bases empiriques de la notion de prédicat sémantique », dans *Langages* n° 63, Paris, Larousse, pp. 7-49.
- Gross (M.), 1986 : *Grammaire transformationnelle du français*, vol. III, *Syntaxe de l'adverbe*, Paris, Asstril.
- Gross (M.), 1990 : « Sur la notion harrissienne de transformation et son application au français », *Langages*, n° 99, Paris, Larousse, pp. 33-56.

- Harris, (Z.), 1976 : *Notes du cours de syntaxe*. trad. Maurice Gross, Paris, Seuil.
- Lerat (P.), 1995 : *Les Langues spécialisées*, Paris, PUF.
- Palma (C.), (travail en cours) : *Estudo contrastivo português-espanhol de advérbios compostos*, Faro, Univ. Algarve.
- Ranchhod (E.), Mota (C.), Baptista (J.), 1999 : « A Computational Lexicon of Portuguese for Automatic Text Parsing », dans *Siglex '99 : Standardizing Lexical Resources* (June 21-22, 1999, University of Maryland, College Park, Maryland, USA), Maryland, University of Maryland. pp. 74-80.
- Silberztein (M.), 1993 : *Dictionnaires électroniques et analyse automatique de textes : le système Intex*, Paris, Masson.
- Silberztein (M.), 2000 : *INTEX Manual*, Besançon, Univ. Franche-Comté.
- Seco *et al.* (2004) : *Diccionario fraseológico documentado del español actual*, Aguilar lexicografía.

DES TERMES AU CONCEPT : LE CAS DU VOCABULAIRE ARCHITECTURAL

Jean-Yves Blaise, Iwona Dudek
UMR CNRS/MCC 694 MAP
Marseille, France
jyb@gamsau.map.archi.fr
idu@gamsau.map.archi.fr

Résumé : *Nous utilisons ici la terminologie comme un outil pour comprendre le bâti patrimonial et ses évolutions. Elle intervient d'abord dans la construction d'un modèle discret, puis, face aux difficultés inhérentes au champ de l'architecture patrimoniale, (réalités difficiles à attacher à un modèle normatif), comme premier filtre d'un processus d'acquisition de connaissances. En réponse, nous présentons un dictionnaire dans lequel chaque item rassemble autour de « l'idée encore floue d'un objet » ses diverses manifestations concrètes.*

Mots-clés : *architecture patrimoniale, analyse terminologique, représentation des connaissances, dictionnaire*

1. INTRODUCTION ET PROBLÉMATIQUE

Dans le champ de l'architecture patrimoniale, l'étude et la compréhension de l'édifice à différentes échelles (du décor aux ensembles monumentaux) s'appuient en priorité sur une analyse approfondie et pluridisciplinaire de sources documentaires variées. Cette analyse a pour vocation, d'une part, d'isoler, dans les mots de J. Cuisenier (1991), « les moments élémentaires dont toute architecture est faite », et d'autre part d'explicitier les vecteurs de transformation du bâti. Elle se heurte en pratique à trois difficultés majeures : l'hétérogénéité des intrants au processus d'analyse (archives, observations contemporaines, connaissances génériques), l'étroitesse (et souvent l'incompatibilité) des champs disciplinaires producteurs d'intrants (histoire, conservation, archéologie, etc.) et enfin le caractère même des intrants (une documentation le plus souvent incomplète, imprécise, incertaine).

Avec le développement des NTIC, la question de l'intégration de ces intrants autour de ce qu'ils contribuent à nous faire comprendre, c'est-à-dire l'évolution d'un lieu architectural, se trouve renouvelée. Sur quels formalismes l'analyste peut-il aujourd'hui s'appuyer pour structurer une étude pluridisciplinaire d'un lieu architectural, en restant fidèle aux inconnues inhérentes à une « science historique » au sens de J. Bocheński (1992, pp. 106-108) ? Cette question, problématique centrale de notre travail, n'est que

très peu abordée comme telle à ce jour¹. Pourtant, les intrants susmentionnés gravitent bien autour d'un lieu architectural, il est par conséquent tentant de tirer profit du caractère spatial des objets manipulés pour les filtrer, les organiser et les visualiser au sens de R. Spence (2001 : 15).

En partant de cette hypothèse, penchons-nous à nouveau sur ce que sont les sources à partir desquelles une analyse de l'édifice peut être menée. Ces sources, documents d'archives textuels ou graphiques, observations contemporaines (*i.e.* relevé), connaissances génériques (*i.e.* théorie architecturale), ont en commun de désigner un lieu architectural. Par là, elles ont en commun des termes, qui peuvent jouer le rôle d'instruments médiateurs de l'analyse. Cependant, ces termes, ces jeux de termes, sont à l'image des édifices eux-mêmes : ambigus, incomplets, multilingues, etc. (figure 1).

Il nous faut en conséquence croiser deux types d'interrogations :

- Quelles sont les normes et règles en vertu desquelles l'édifice est conceptualisé puis bâti, et par conséquent quelle est la terminologie qui le décrit ?
- Quelles sont les logiques de transformation de l'édifice, et par conséquent quelles décontextualisations terminologiques impliquent-elles ?

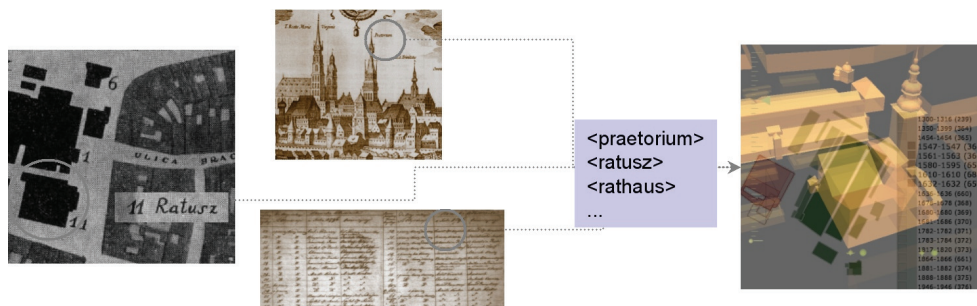


Figure 4: trois vocables différents, issus de trois sources documentaires différentes, pour repérer un même lieu

Ce double questionnement peut être d'une certaine façon rapproché de l'opposition synchronie/diachronie qu'établissait F. de Saussure. L'effort de description du lieu nécessite la construction d'un modèle discret permettant d'identifier un jeu de concepts fondamentaux, concepts reconnus par un jeu de termes, aptes à être évalués en matière de contexte d'utilisation. Les vocables repérant ces concepts architecturaux deviennent dès lors une grille d'analyse de l'édifice, rapportant la forme de celui-ci à une connaissance générique, et permettant de comparer désignation *a priori* du vocable et utilisation spécifique, contextuelle.

¹ En effet, de nombreux travaux sont menés dans le cadre de tel ou tel champ disciplinaire, comme par exemple la gestion documentaire avec le travail de R. Stenvert (1991), l'acquisition de données spatiales et ses problèmes d'exploitation (voir Ramondino (2001)), ou encore l'informatique graphique et ses applications discutables (et discutées dans Kantner (2000)). Mais l'étape consistant à intégrer les résultats de ces travaux autour de leur sujet, un lieu, n'est pour l'essentiel pas franchie.

Le concept architectural devient une racine univoque permettant de rassembler autour de « l'idée de la chose » ses diverses manifestations concrètes, termes et expressions exprimés dans une langue donnée, dérivations locales de ces termes, interprétations régionales de la forme théorique que le concept identifie, etc. Cette approche pose deux grandes familles de problèmes :

- Quelle méthode pour identifier ces concepts de façon univoque, quel modèle discret mettre en œuvre ?
- Quel moyen, quels outils concrets pour mettre en partage le vocabulaire désignant ces concepts, et rapporter au dit vocabulaire ses déclinaisons contextuelles ?

En réponse, notre première contribution est l'élaboration d'un modèle architectural qui définit des concepts identifiés par filtrage du vocabulaire puis organisés dans une structure hiérarchique. Mais cette première contribution ne répond qu'à une de nos préoccupations, celle d'identifier des concepts architecturaux de façon univoque. Reste alors à autoriser l'étude de leurs déclinaisons contextuelles. Il faut ici noter que ce modèle, déterministe, dit « intensionnel », ne recouvre que partiellement la réalité souvent confuse du bâti patrimonial : le temps s'est chargé de transformer largement les objets qui nous intéressent, à tel point que les réalités que nous observons aujourd'hui sont souvent en rupture avec les définitions canoniques des concepts architecturaux mentionnées plus haut. Bien souvent, l'objet n'est plus ce qu'il devrait être : ni sa forme ni son rôle ne perdurent (figure 2).

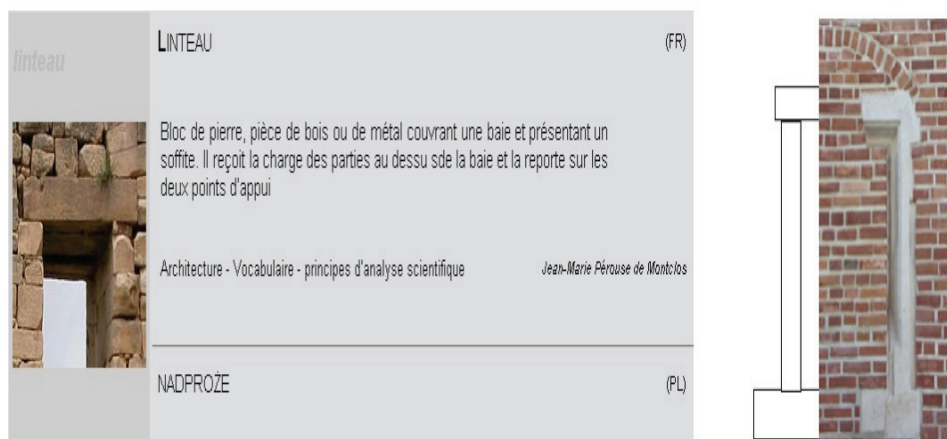


Figure 5: À gauche, identification du concept « linteau » et une de ses définitions ; à droite, l'objet figuré est-il reconnaissable comme une baie couverte par un linteau ? Le fragment de linteau préservé relève-t-il de la notion de linteau ?

Seul le terme par lequel nous pouvons encore le désigner nous permet de rapporter l'objet à un concept. La terminologie devient alors le moyen de reconnaître l'objet, et

par conséquent le seul descripteur réellement utilisable en phase amont de l'étude. Dès lors, l'identification du vocable permettant de désigner la chose observée devient une première étape dans un processus d'acquisition d'informations dont la dernière étape, celle qui consiste à rapporter l'objet à un élément du modèle déterministe mentionné plus haut, ne sera pas franchie par tous (voir figure 2).

Pour la mise en œuvre concrète de la première étape de ce processus, nous développons un outil appelé *Diva*, dictionnaire méthodologique permettant de lier un concept architectural univoque, représentable spatialement, un jeu de vocables relevant d'une analyse terminologique générique, construction du sens, et enfin un jeu de contextes utilisés à fins de comparaisons.

Dans cette contribution, nous poserons brièvement les hypothèses à partir desquelles ce travail est mené, puis nous développerons trois points principaux : identification d'un *corpus* univoque, limites de cette première approche, et enfin rôle, principes et implémentation de l'outil terminologique *Diva* construit en réponse.

2. HYPOTHÈSES

Nous essayons de construire des représentations abstraites de réalités qui ont existé ou existent encore pour :

1. comprendre et retracer les évolutions de lieux architecturaux à travers l'histoire ;
2. organiser et donner accès aux jeux de données et d'informations ayant permis l'étude des dites évolutions.

Pour ce faire nous partons de trois hypothèses :

- le *corpus* architectural, les formes physiques élémentaires constituant l'édifice, sont un médiateur entre les différents jeux de données et d'informations ;
- puisque travaillant sur des objets disparus ou transformés, ces formes seront toujours mal connues, et représentées de façon abstraite ;
- dans la maquette 2D ou 3D, ces formes sont un outil de navigation puisqu'elles permettent à l'utilisateur d'interroger élément par élément les jeux de données et d'informations organisées en conséquence.

C'est donc bien l'idée d'un lieu bâti qui est présente, et non une réalité physique donnée, et c'est cette idée d'un lieu bâti qui est utilisée comme fédérateur. La terminologie, comme on le verra, est au cœur de cette démarche, même si elle se retire avec discrétion de l'expression spatiale ci-après.

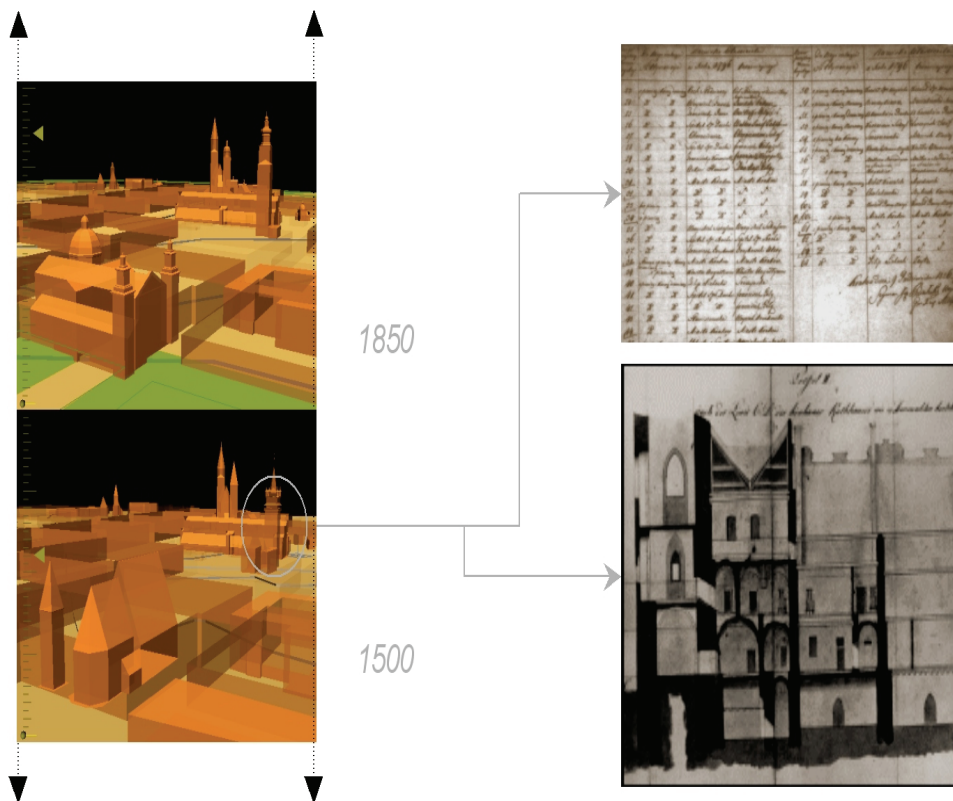


Figure 6 : des maquettes virtuelles, à caractère symbolique, utilisées comme moyen de tri et d'accès à des jeux de documents (dans l'exemple, deux documents attachés à l'objet entouré, l'ancien hôtel de ville de Cracovie)

En effet, dans ce travail nous partons de trois familles de sources. Entre ces différentes sources et l'idée fédératrice et abstraite d'un lieu bâti, que nous recherchons, il y a un médiateur naturel, des termes. D'où le titre de cette contribution, et l'idée simple de se servir du concept pour relier, avec les différents termes qui le désignent, les différentes sources qui le documentent.

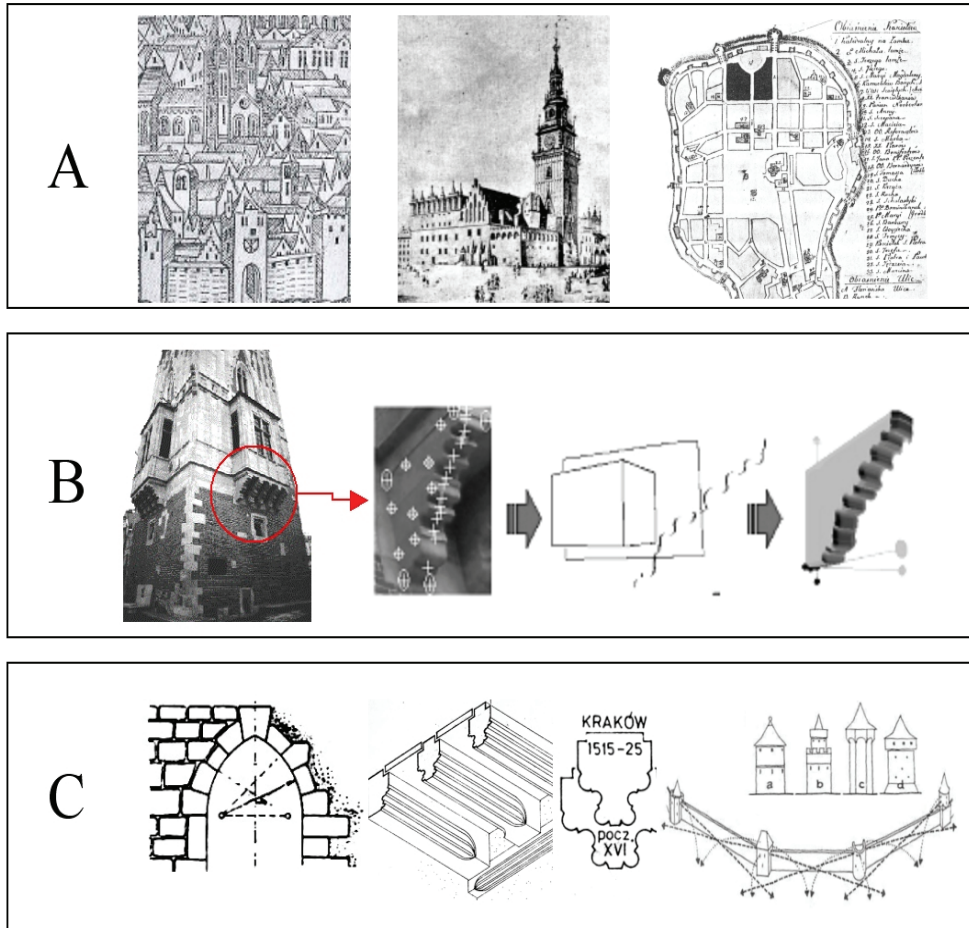


Figure 7 : trois grandes familles de sources documentaires : A, archives, B, observations, C, théories

3. IDENTIFICATION DES ÉLÉMENTS DU *CORPUS* : UN MODÈLE DISCRET

Notre objectif est de tirer profit du caractère *spatial* des objets manipulés pour filtrer, organiser, visualiser les jeux de données attachés aux édifices au travers de maquettes virtuelles 2D/3D. La question que nous devons nous poser est en conséquence la suivante : à quel concept spatial décrivant l'édifice peut-on attacher les données que la maquette interface ? Notre premier problème a donc été d'isoler d'une manière univoque des éléments significatifs du *corpus* architectural, et donc d'établir une méthode globale d'analyse et de description de l'édifice.

3.1. Principes de constitution du modèle

Nous regardons le lieu bâti comme une combinatoire d'objets, relevant chacun d'une échelle donnée², dont les propriétés décrivent morphologie, position et fonctions. La classification des éléments proposée revient à construire une hiérarchie de concepts (classes au sens de la programmation orientée objet) et à formaliser les relations qui les lient. Nous nous appuyons sur des modèles canoniques d'architecture, repérés par des termes, dans un raisonnement de type intensionnel. L'édifice est décrit par l'observation de régularités qui tiennent autant à des typologies qu'à des observations structurelles. Dans ce travail d'identification de concepts, le langage architectural est l'outil servant à « discrétiser » l'édifice.

Ce passage « du terme au concept », présenté dans Blaise (2004 : 574-577), peut s'analyser à travers la grille que propose F. Rastier (1995 : 35-65) et qui distingue quatre étapes :

(i) « La nominalisation donne pour forme canonique du terme le nom substantif [...]. La nominalisation est fort utilisée pour créer un effet d'objectivation [...].

(ii) La lemmatisation permet ensuite, outre sa commodité lexicographique, de dépouiller de ses variations accidentelles la substance que le terme est censé représenter [...].

(iii) La décontextualisation permet de définir le terme par lui-même, indépendamment des variations qui pourraient affecter ses occurrences [...].

(iv) La constitution du mot en type, et l'affirmation corrélatrice que toutes ses occurrences sont subsumées sous ce type – ou du moins que celles qui ne le sont point témoignent d'un emploi incorrect. La définition est le moyen principal de cette constitution en type : elle énonce, conformément au principe du positivisme logique, les conditions nécessaires et suffisantes pour que le terme soit pourvu de sa dénotation correcte. »

L'élaboration du *corpus* à partir des apports lexicographiques dont nous nous servons se fait par filtrage des termes rencontrés pour en extraire des concepts univoques et significatifs du point de vue du langage architectural³. À partir des lexicographies, qui permettent d'isoler des formes canoniques (la soliveAProfil), notre travail consiste à défaire chaque terme ou expression des déclinaisons morphologiques du terme (la soliveAProfil = une solive + un profil, le Profil = un jeu de moulures renvoyant à une typologie de la mouluration). Une autre étape consiste alors à sortir l'élément que désigne le terme (représentant le concept que nous cherchons à isoler) des conditions matérielles de son emploi (la soliveAncrée = une solive + un ancrage, l'arcature aveugle = un jeu d'arcs et de remplissages).

² On entend ici « échelle » non comme rapport dimensionnel mais au sens de niveaux de conception établi par P. Boudon (1977).

³ Le problème d'identification des concepts architecturaux auquel nous sommes confrontés croise ainsi celui de la terminologie dans la définition qu'en donne H. Felber (1987 : 1) « Domaine du savoir interdisciplinaire et transdisciplinaire ayant trait aux notions et à leurs représentations » cité par F. Rastier (1995 : 35-65)

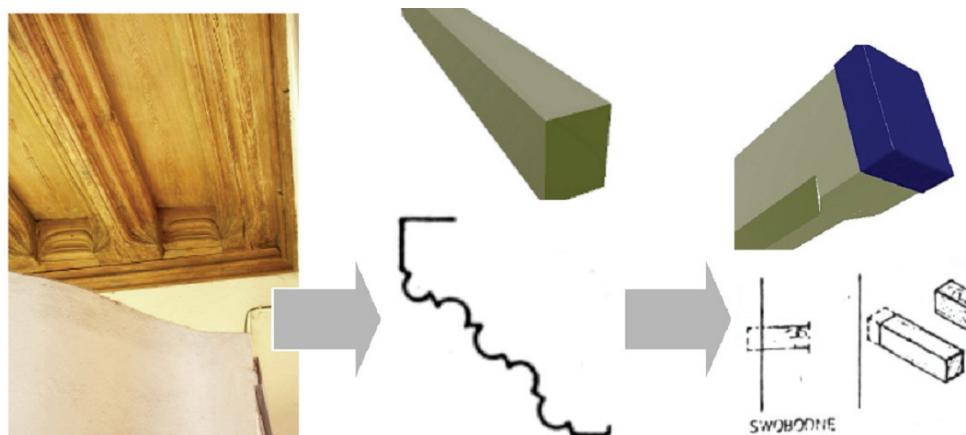


Figure 8 : Trois étapes de l'analyse du vocabulaire : formes canoniques, déclinaisons morphologiques, conditions matérielles d'emploi

Le modèle résultant fixe un canevas de règles permettant d'isoler au sein du *corpus* correspondant à chaque expérimentation les concepts architecturaux à identifier, nommer et structurer. Ce canevas constitue de fait une méthodologie de relecture des références. Y compris dans son implémentation, il s'appuie sur les principes fondamentaux et le pragmatisme de *l'approche-objet* tels que décrits par R. Ducournau (1998). Chaque concept définit un moule servant à construire des éléments particuliers. Le concept décrit une forme architecturale par les descripteurs qui la caractérisent (par exemple, centre, rayon et profondeur d'un arc plein cintre), et chaque instance est une forme architecturale caractérisée par des valeurs particulières données à chacun des descripteurs. Les concepts ainsi isolés sont organisés hiérarchiquement par l'observation de similitudes morphologiques ou fonctionnelles.

4. VERS L'IMPLÉMENTATION D'UN PROCESSUS D'ACQUISITION D'INFORMATIONS

Le modèle présenté ci-dessus possède des qualités de robustesse et de cohérence vis-à-vis du formalisme informatique confirmées par exemple dans Blaise (2003b : 128-134). Mais dans le champ de l'architecture patrimoniale, l'analyste se heurte à des réalités bien moins régulières que ne l'est ce modèle intensionnel, fondamentalement déterministe.

Un exemple est présenté ci-dessous (figure de gauche) : la rive d'une couverture d'église, attribut d'un objet de type versant de toit, se retrouve imbriquée dans un élément de type mur. Le mur est-il devenu un versant de toit ? La rive un attribut de mur ? La couverture une partie du mur ? Le mur ou la rive sont-ils devenus des sortes de couvertures d'églises ?

La structure cognitive commodément régulière que nous avons mise en place chancelle. Il faut observer que, parce que nous utilisons la morphologie de l'objet comme

ligne de division et de classification du modèle, la détermination morphologique du concept peut ne recouvrir que partiellement celle de l'instance. On le voit sur cette même figure, en partie droite : l'objet originel (une voûte en berceau, comme en témoigne l'inclinaison des murs) a perdu non seulement sa fonction mais aussi sa morphologie (construite en théorie sur un profil en demi-cercle continu, ce profil est séparé dans l'exemple en deux portions de cercles disjointes).

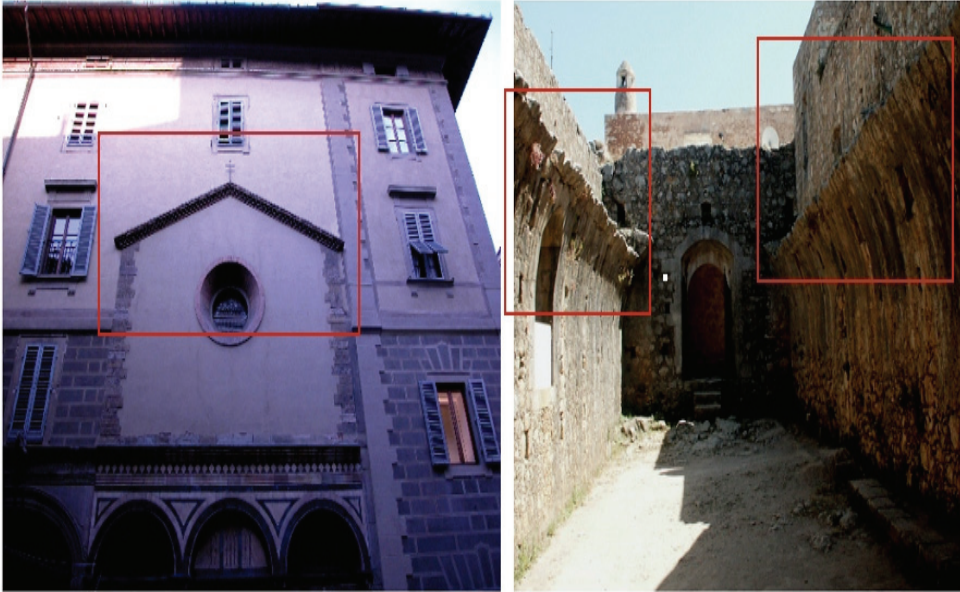


Figure 9 : des objets qui ne répondent plus à leur définition normative

Notre modèle déterministe, dit intensionnel, est efficace du point de vue informatique, mais pose toute une série de problèmes brièvement illustrés par l'exemple, dans la figure ci-après.



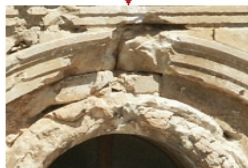
L'instance n'est pas identifiable comme relevant d'un élément du modèle :

- gauche, propriétés manquantes (la baie n'en a plus ici les propriétés d'ouverture)
- centre, objet non reproduits
- droite, objets altérés ou fragmentaires



L'instance a, par rapport à l'élément du modèle dont elle relève :

- des propriétés supplémentaires « accidentelles », le plus souvent uniques (altérations de l'architrave)
- des propriétés aux valeurs incohérentes (longueurs relatives du linteau et de son larmier)



L'instance ne joue plus le rôle qui lui incombe dans le système de relations de l'édifice (ombre portée coupée par l'altération de l'archivolte)

Figure 10 : Objets réels manifestant les limites d'un modèle déterministe

Ce qu'il faut retenir c'est qu'il est souvent difficile de dire d'un objet lambda, surtout au début de son étude : « cet objet est de type X avec telle et telle propriété ». Nous avons donc évalué l'efficacité du modèle discret en nous disant qu'il nous faut regarder l'objet réel non pas avec pour objectif de le faire rentrer dans une case du modèle déterministe, mais avec pour objectif de gratter couche par couche les éléments d'informations pour *in fine*, si possible seulement, instancier ce modèle. Autrement dit, il nous faut implémenter les phases d'un processus d'acquisition de connaissances non normatif, adapté aux réalités de la pratique. Le schéma suivant en montre la logique, que nous ne détaillons pas plus ici. Mais on notera que la première étape de ce processus fait encore intervenir le vocabulaire, cette fois-ci comme premier filtre au travers

duquel on va tenter de comprendre un objet réel et le mettre en relation avec l'univers de connaissance dont il relève.

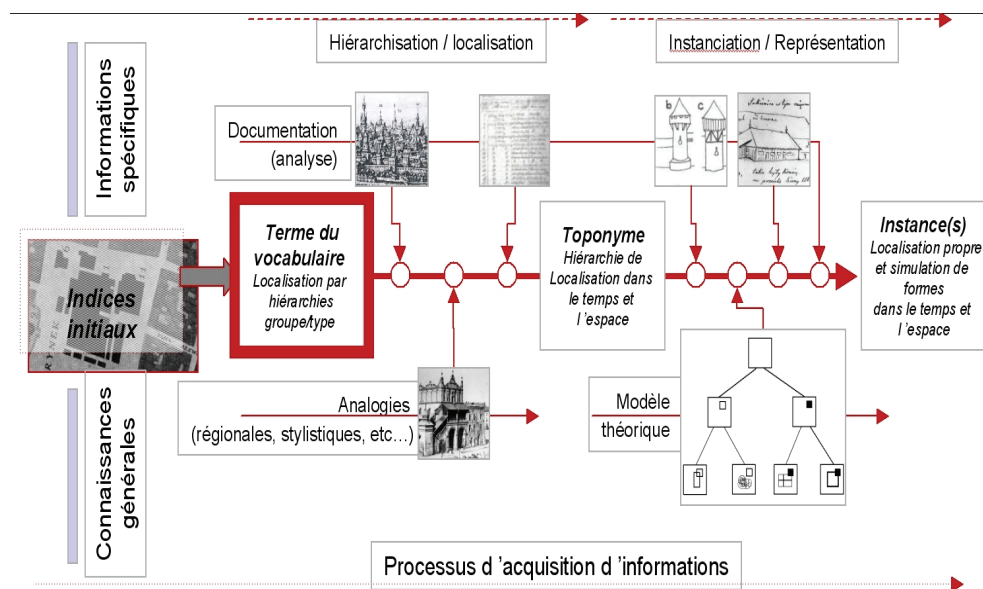


Figure 11 : Un processus de recueil d'indices en phase amont de l'étude

Le dictionnaire *Diva*, dont nous introduisons les principes dans la section suivante, doit être ainsi compris comme l'outil implémenté pour instrumenter la première étape d'un processus d'acquisition de connaissances, encadrée de rouge ci-dessus.

5. LE DICTIONNAIRE *DIVA*

Puisque nous en avons situé dans la section précédente le rôle, nous présenterons ici le dictionnaire méthodologique *Diva* comme une application autonome. *Diva* est l'acronyme de *dictionnaire méthodologique pour le vocabulaire architectural*. *Diva* a pour premier objectif d'assister l'identification et la mise en comparaison d'un concept architectural, des termes et des réalités qui le désignent. Il a pour second objectif de jouer le rôle d'un lexique multilingue illustré. *Diva* est construit autour de cinq principes que nous allons rapidement fixer.

Principe 1 : distinguer le concept et les termes

Diva implémente la distinction concept/termes, ou signifié/signifiants, notion déjà en filigrane du travail d'identification des concepts présenté en section 3. Un item du dictionnaire c'est « l'idée de la chose », un concept architectural, auquel plusieurs termes peuvent être attachés. Il n'y a donc pas de traductions terme à terme mais la mise en relation de termes avec un concept unique. Par là, cette approche allie comparaisons de

contextes et rapport à un référent commun, comme on le voit dans les deux dérivations contextuelles du concept d'arc brisé ci-dessous. On remarquera, en bas à droite, que l'arc brisé est rebaptisé par J. Le Goff « Arc-doubleau » en référence à son rôle dans le voûtement décrit.

Autrement dit, l'arc brisé, dont la définition univoque est symbolisée par le graphique de gauche, est rebaptisé eu égard à un contexte d'utilisation (la « croisée d'ogives ») que l'auteur veut expliquer.

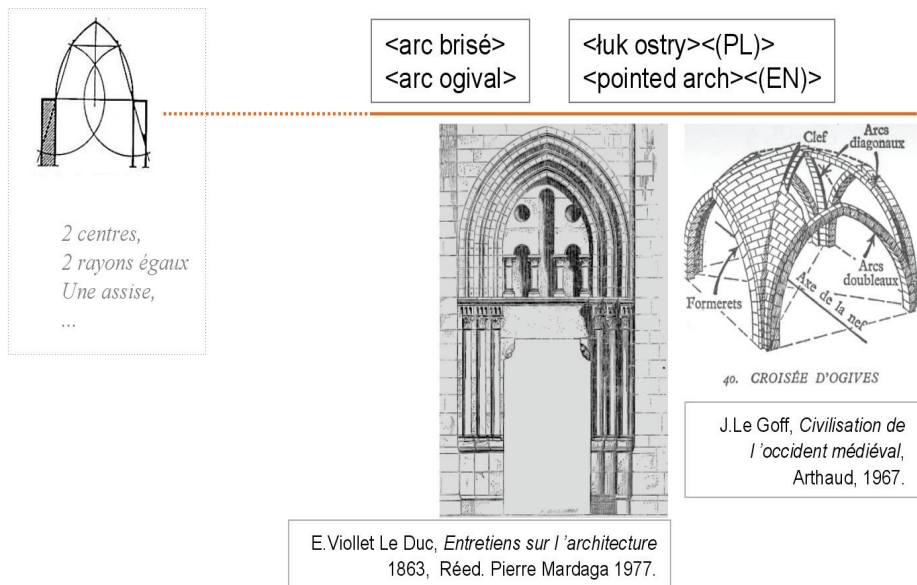


Figure 12 : Le concept « arc brisé » et ses déclinaisons

Principe 2 : Mult-isme

Diva est un lexique multidéfinitions et multilingue : à un *item* du dictionnaire sont rapportées un nombre libre de définitions (multilingues) et de traductions, permettant d'enrichir le *corpus* sur le long terme. Une analyse des termes relevés dans chaque définition permet de renvoyer vers d'autres items du dictionnaire (encadrés de gauche sur la capture d'écran du système reproduite plus loin). La mise à jour et la consultation se font à distance en utilisant la toile et des formats dits « libres ».

Principe 3 : Tout référencer

Les définitions et traductions proposées ne sont pas notre contribution : nous nous attachons à utiliser des travaux de références et à les citer. À chaque définition ou traduction correspond donc une référence bibliographique affichée systématiquement à côté

du texte ou du mot (encadrés de droite sur la capture d'écran du système reproduite ci-dessous).

Ce choix permet de comparer des propositions entre elles, et par rapport à cette « idée de la chose » nous tentons de circonscrire.

PODIUM

Reference of theme (1- P) :
Architecture - Vocabulaire - principes d'analyse scientifique (FR)
 Imprimerie nationale, Paris, 1988
 Jean-Marie-Pérouse de Montclos (1)

PODIUM: (FR)
 Haut soubassement avec un ou plusieurs degrés d'accès (voir degré).

degré- soubassement: (FR)
Reference (1- P) :
Architecture - Vocabulaire - principes d'analyse scientifique (FR)
 Imprimerie nationale, Paris, 1988
 Jean-Marie-Pérouse de Montclos (1)

PODIUM: (FR)
 Massif élevé au-dessus du sol, comportant normalement des parois verticales, et supportant un (ou plusieurs) édifice(s). Le PODIUM, normalement SIMPLE, peut être DOUBLE quand il comporte deux massifs superposés, le plus haut parfois de dimensions réduites. Le plan supérieur du PODIUM est rendu accessible par un ESCALIER. On trouve normalement, de bas en haut, les éléments suivants : BASE, CORPS, COURONNEMENT.

base- podium- sol- (FR)
Reference (3- G) :
Dictionnaire méthodique de architecture grecque et romaine (fr)
 Tom 1 - II Ecole Française d'Athènes
 Ecole Française de Rome, 1998
 René-Ginouvès (3) R-Martin (4)

PODIUM, PODEST: (DE)
Reference(s) of Translation (3- G) :
Dictionnaire méthodique de architecture grecque et romaine (fr)
 Tom 1 - II Ecole Française d'Athènes
 Ecole Française de Rome, 1998
 René-Ginouvès (3) R-Martin (4)

PODIO: (IT)
Reference(s) of Translation (3- G) :
Dictionnaire méthodique de architecture grecque et romaine (fr)
 Tom 1 - II Ecole Française d'Athènes
 Ecole Française de Rome, 1998
 René-Ginouvès (3) R-Martin (4)

Figure 13 : Application concrète des principes 2 et 3

Principe 4 : Relations = navigation

Les relations sémantiques entre items du dictionnaire sont utilisées comme mode de navigation au travers de liens Internet simples (URL). Parmi ces relations, deux types de relations relevant de la sémantique architecturale sont identifiées :

- relation partie-de ou *groupe parent* (renvoie à la notion d'échelle architecturale) ;
- relation sorte-de ou *objet parent* (renvoie aux propriétés géométriques des objets).

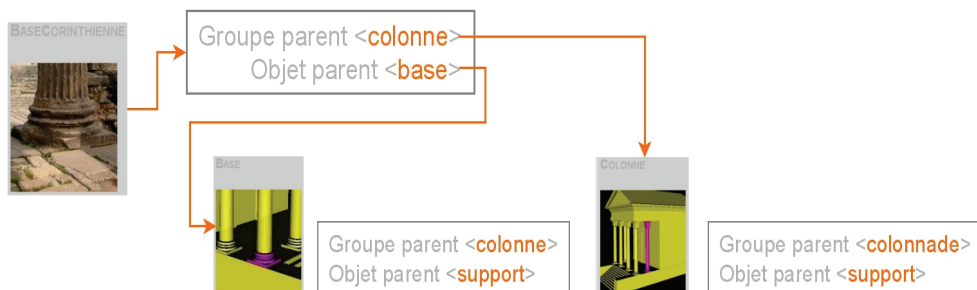


Figure 14 : Les relations groupe parent et objet parent illustrées sur le cas de la base corinthienne

Ces relations créent des liens de dépendance entre items qui doivent permettre de poser une question « par raffinements successifs »⁴.

D'autres relations sont également utilisées, qui s'appuient sur le vocabulaire lui-même :

- La relation de « citation » (analyse des termes relevés dans chaque définition) ;
- La relation d'exclusion, qui relève de la notion d'antonyme développée par J.-M Pérouse de Montclos (« ne pas confondre A avec B car... ») ;
- Une relation de groupe typologique qui s'apparente aux qualifieurs de groupe de ce même auteur.

5.1 Illustrer le concept par des objets particuliers

Un item du dictionnaire permet d'attacher plusieurs termes à un concept, mais aussi plusieurs déclinaisons contextuelles de ce concept. Le mécanisme en est simple. Au départ, nous avons d'un côté un concept repéré par des termes, de l'autre un lieu architectural donné. Ce lieu est décrit par divers documents qui explicitement ou implicitement renvoient à un ou plusieurs concepts. Si l'on répète l'opération sur plusieurs lieux, on a bien une relation circulaire entre concept, lieux et documents.

⁴ Par exemple : quel est le nom de la partie cylindrique d'une colonne ? Voir concept colonne, relation partie de [réponse : fût]. – quel est le nom donné aux fûts aux profils en concavités inversées ? Voir concept fût, relation sorte de [réponse : fût rudenté].

Chaque ressource documentaire est mise en relation avec le terme général qu'elle illustre et avec l'édifice particulier qu'elle documente.

La terminologie structure les relations entre concept, objets particuliers et ressources documentaires

- Dans l'espace des concepts (la ressource R illustre le vocable V) ;
- Dans celui des instances (la ressource R documente l'objet O, matérialisation du vocable V).

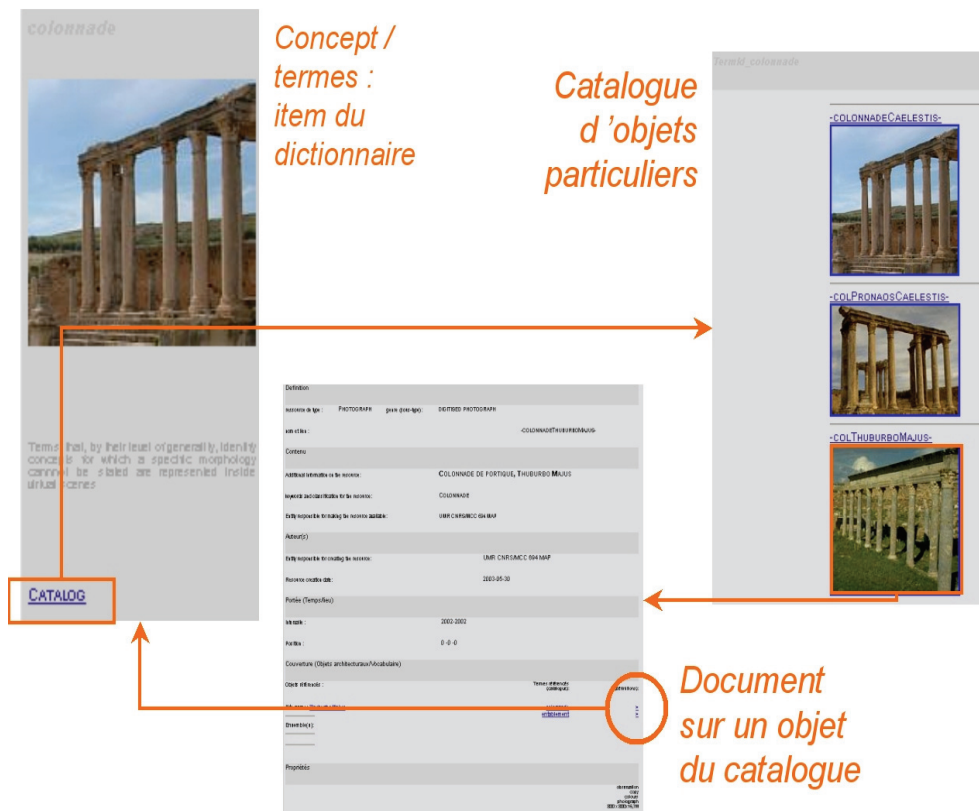


Figure 15 : relations en triangle concept/objets/documents, cas de la colonnade

5.2 Implémentation et portée

Diva est une application développée pour le réseau Internet à partir de formalismes dits libres uniquement. Les documents individuels, items du dictionnaire, références, ressources, etc., sont des fichiers XML. En consultation, ils sont affichés par un navigateur standard pour Internet après interprétation de feuilles de style XSLT. Les modifications sont réalisées par le biais de scripts Perl produisant les formulaires XHTML adéquats. L'interface de consultation actuelle est intégrée dans un navigateur

d'applications qui permet, à partir d'un lieu architectural donné, d'en balayer les différents déterminants : vocabulaire bien sûr, mais aussi toponymie, typologie, représentations spatiales 2D/3D, etc.

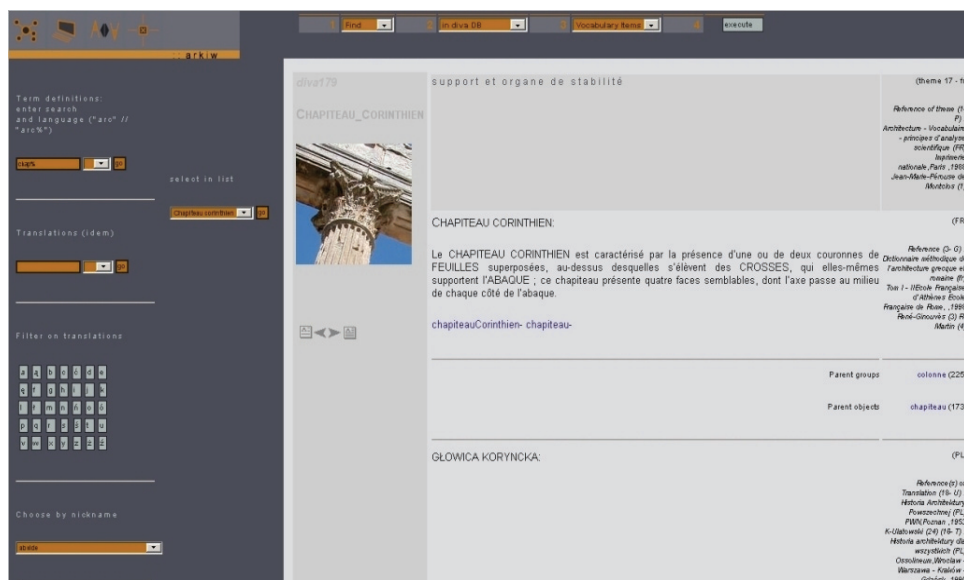


Figure 16 : L'interface de consultation ; en partie droite (fond gris) tri thématique, définitions, traductions, renvois vers la bibliographie, relations parent/objet/citation ; en partie gauche outils de sélection et de requêtes, en partie haute liens vers les autres applications

Diva est une application modeste, expérimentale, qui recense aujourd'hui à peu près 800 termes en 7 langues. Il a été testé dans le cadre d'un programme européen, le programme *Strabon*, dans lequel nous avons pour mission d'étudier l'interfaçage 2D/3D de jeux de données à caractère culturel.

Notre réponse a été d'organiser les jeux de données en les rattachant à des items terminologiques au regard des principes évoqués ci-avant, et d'utiliser la représentation, dans l'espace de maquettes virtuelles 2D/3D à différentes échelles, d'objets particuliers correspondant à ces items⁵.

⁵ Un site de démonstration rudimentaire est accessible à l'adresse <http://www.map.archi.fr/strabon>

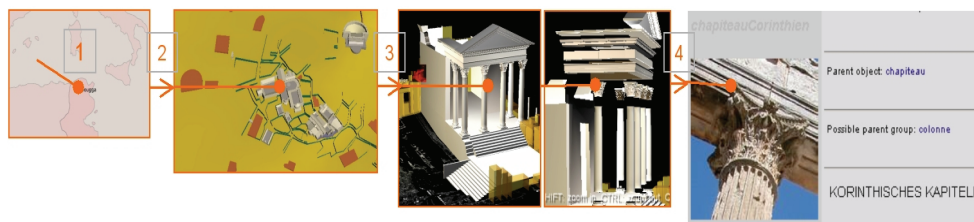


Figure 17 : Imbrications d'échelles utilisées comme filtre pour interroger, soit les items terminologiques, soit les jeux de données (une même maquette permet d'interroger soit l'édifice particulier présenté, ici le capitole de Douggha, soit les concepts dont l'édifice relève, c'est-à-dire le dictionnaire)

6. CONCLUSION

En résumé, dans notre expérience, la terminologie s'est d'abord avérée comme un outil efficace pour construire un modèle discret du lieu bâti. Mais parce que nous travaillons dans le champ de l'architecture patrimoniale, nous analysons des réalités construites souvent confuses. En conséquence, il nous faut implémenter les phases d'un processus d'acquisition de connaissances, processus dans lequel la terminologie intervient à nouveau, cette fois-ci comme premier indice.

Dans ce cadre, nous avons construit un dictionnaire méthodologique qui reste à améliorer sur bien des points (par exemple, sur l'extension de la notion de relation aux synonymes, noms anciens, etc. ; mais aussi sur la représentation graphique des relations d'un concept ou sur les modalités de recherche).

Plus généralement, parce qu'il donne aux termes ce rôle de point commun dans des jeux d'informations, notre travail nous semble sortir du strict champ de l'architecture patrimoniale, et nous pousse à investiguer de possibles passerelles interdisciplinaires avec la communauté LTT, notamment afin d'évaluer les conditions d'un développement plus conséquent de la plate-forme *Diva*.

BIBLIOGRAPHIE

- Alkhoven (P.), 1993 : *The changing image of the city. A study of the transformation of the townscape using Computer assisted Design and visualisation techniques*, thèse de doctorat, Université d'Utrecht (Pays-Bas).
- Blaise (J.-Y), 2003a : *Le Formalisme objet appliqué à l'étude de l'architecture patrimoniale – problèmes de modélisation et d'échanges de données sur le réseau Internet*, thèse de doctorat, Université Aix-Marseille III, mars 2003.
- Blaise (J.-Y) et Dudek (I.), 2003b : « Exploiting the architectural heritage's documentation : a case study on data analysis and visualisation », *Actes conférence I-Know 03*, Graz (Autriche), Journal of Universal Computer Science.

- Blaise (J.-Y) et Dudek (I.), 2004 : « Règles d'identification et méthodes de visualisation d'objets architecturaux », *actes de la conférence : « Extraction et Gestion de connaissances »*, EGC 04, Clermont-Ferrand, France.
- Bocheński, (J. M.), 1992 : *Współczesne metody myślenia* ; W Drodze/Poznań.
- Boudon (P.), 1977 : *Richelieu ville nouvelle*, Dunod Éditeur.
- Cuisenier (J.), 1991 : *La Maison rustique ; logique sociale et composition architecturale*, PUF.
- Dmochowski (Z.), 1956 : *The architecture of Poland, Polish research centre*, London/Alma Book.
- Ducourmau (R.), Euzenat (J.), Masine (G.), Napoli (A.), 1998 : *Langages et modèles à objets, état des recherches et perspectives*, publié par l'INRIA (Institut national de la recherche en informatique et en automatique).
- Felber, H. (1987) *Manuel de terminologie*, UNESCO, INFOTERM.
- Kantner (J.), 2000 : « Realism vs Reality : creating virtual reconstructions of prehistoric architecture », dans J.A. Barcelo, M. Forte, D. H. Sanders (édit.) *Virtual reality in archaeology*, Oxford, Archeopress.
- Pérouse de Montclos (J. M.), 1988 : *Architecture vocabulaire-Principe d'analyse scientifique*, Imprimerie Nationale.
- Ramondino (F.), 2001 : *From point cloud to surface : the modeling and visualisation problem – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXIV- 5/W10.
- Rastier (F.), 1995 : Le terme : entre ontologie et linguistique, *La banque des mots*, n° 7.
- Spence (R.), 2001 : *Information vizualisation*, Addison Wesley ACM Press.
- Stenvert (R.), 1991 : *Constructing the past : computer-assisted Architectural-Historical Research*, thèse de doctorat, Université d'Utrecht (Pays-Bas).

CONTEXTE D'UTILISATION, CONTEXTE DE COMMUNICATION – LA DOUBLE IDENTITÉ DU DICTIONNAIRE D'ENTREPRISE

Patrick Leroyer

Institut de langue et de communication d'entreprise
section de français, centre de lexicographie
École des hautes études commerciales d'Århus
Århus, Danemark

Henrik Køhler Simonsen

École des hautes études commerciales de Copenhague
département d'anglais
Frederiksberg, Danemark

Résumé : *la conception du dictionnaire d'entreprise ne peut simplement reposer sur une démarche terminologique classique, mais doit inclure une démarche centrée sur les besoins cognitifs et communicationnels des utilisateurs, c'est-à-dire le contexte d'utilisation, et sur le plan de communication de l'entreprise, c'est-à-dire le contexte de communication. Nous jetterons ici les bases d'une métalexicographie d'entreprise fonctionnelle, et présenterons quelques exemples de solutions lexicographiques et technologiques offrant, grâce aux genres textuels, un accès dynamique aux termes et à leurs contextes.*

Mots-clés : *dictionnaire d'entreprise, fonctions lexicographiques, genres textuels, contextes dynamiques*

1. INTRODUCTION : ONTOGÉNÈSE DU DICTIONNAIRE D'ENTREPRISE

Les textes spécialisés dans l'entreprise ne relèvent pas seulement du ou des domaines de l'entreprise, mais sont déterminés par l'image de l'entreprise et par toutes les facettes de son métier, et par la multiplicité des points de vue de ses différents locuteurs à l'interne comme à l'externe. Ces textes servent d'instanciations communicatives dans des situations et autour de problèmes de communication spécifiques, et participent simultanément à la construction, à la représentation et à la communication des connaissances de l'entreprise, tandis que leur genèse et leurs modalités d'utilisation restent soumises au management stratégique et au plan de communication de l'entreprise. Ainsi, les textes de la communication *corporate* (relations publiques, communication interne, marketing, responsable, financière, de crise, etc.) sont-ils à inclure dans le champ de ces textes spécialisés. Vecteurs de connaissances selon Lerat (1995), ils sont aussi vecteurs de choix communicatifs valorisants et de messages spécifiques qui les enrichissent de sens. Pour une entreprise fabriquant et commercialisant des pompes par

exemple, il va sans dire qu'un dictionnaire d'entreprise (= DE) ne saurait se limiter à un dictionnaire des terminologies des pompes, aussi complet et systématique fût-il.

Que faire dès lors pour fournir une aide optimale aux utilisateurs travaillant dans l'entreprise au moyen d'ouvrages de référence lexicographiques performants ? La conception de tels ouvrages passe nécessairement par une transformation des bases terminologiques maison traditionnellement développées par les entreprises ou par les fournisseurs de systèmes de banques de données terminologiques et de gestion des contenus. Pour expliquer et illustrer cette transformation, nous présenterons ici les principaux éléments d'une théorie générale de la lexicographie d'entreprise et de sa pratique.

2. MÉTALEXICOGRAPHIE DU « DE »

Le DE doit être conçu comme un dictionnaire à double contexte de communication et d'utilisation. De communication d'une part, parce que le DE doit enregistrer et représenter systématiquement les textes et les corps de connaissances de l'entreprise, en conformité avec sa politique et son plan de communication. D'utilisation d'autre part, parce que le dictionnaire étant par nature un outil fonctionnel, il doit être conçu pour remplir son objectif canonique qui est de résoudre les problèmes d'utilisateurs spécifiques dans des situations d'utilisation spécifiques, qui travaillent sur des textes – lorsqu'il s'agit de lire, d'écrire ou de traduire – ou sur des connaissances – lorsqu'il s'agit d'apprendre, de comprendre, ou de se documenter. Ce qui est nouveau ici – pour les deux catégories de problèmes – est que l'ensemble des solutions lexicographiques orientées vers les besoins des utilisateurs doivent être développées en conformité avec la politique de communication de l'entreprise. Le dictionnaire doit être bâti comme une base terminologique et textuelle intégrée offrant aux utilisateurs des solutions lexicographiques et politiques aux problèmes s'orientant vers la communication ou vers la connaissance. En ce sens, le DE, qui répond aux besoins de ses utilisateurs individuels tout en restant fidèle à la démarche collective de l'entreprise, est un outil polyfonctionnel à caractère profondément prescriptif, et repose sur les principes axiomatiques de la lexicographie fonctionnelle formulés par Bergenholtz et Tarp (2003), ainsi que Tarp (2001), et attribuant à la lexicographie le statut de science à part entière, l'objet de son étude étant évidemment les dictionnaires.

2.1. Fonctions lexicographiques du DE

Le DE est caractérisé par son contexte d'utilisation, c'est-à-dire les situations d'utilisation et les profils des utilisateurs de l'entreprise. Ces profils complexes se mesurent en termes de compétences, de besoins cognitifs et communicationnels, et en termes de situations d'utilisation du dictionnaire. À ces situations et profils correspondent en effet deux catégories de fonctions et deux scénarios de consultation différents. Les fonctions cognitives d'une part, qui sont orientées vers la connaissance, où l'on consulte le dictionnaire parce qu'il y a quelque chose que l'on ne sait pas, pour connaître le cœur de métier de l'entreprise par exemple, sa politique de communication et sa stratégie, sa gouvernance, sa politique linguistique, ses valeurs fondamentales et ses

messages génériques, mais indépendamment de toute activité textuelle. Les fonctions communicationnelles d'autre part, où l'on consulte le dictionnaire parce qu'il y a problème pendant que l'on travaille sur un texte, pour lire, traduire ou produire un texte, pour l'adapter ou le localiser, ou pour transcrire des messages spécifiques tout en respectant les principes de la politique de communication et de la politique linguistique de l'entreprise, le dictionnaire faisant alors office de simple « relais » dans le processus textuel.

2.2. Lexicographie des genres textuels

Ancré dans un contexte spécifique et déterminé par ses fonctions lexicographiques, le DE doit donc tenir compte du métier de l'entreprise, de sa stratégie, de son plan de communication et de ses enjeux, et de ses parties prenantes. La solution fédératrice que nous avons développée consiste à concevoir un dictionnaire basé sur le traitement systématique des genres textuels privilégiés par l'entreprise. Les genres textuels ont en effet trois qualités prescriptives centrales. La qualité « communication de la connaissance », car les genres textuels actualisent des stratégies de simplification et d'explicitation particulièrement appropriées à la fonction apprentissage des langues spécialisées et des connaissances du métier par exemple. Les qualités « innovation » car les genres textuels actualisent les toutes dernières connaissances et expériences de l'entreprise. Les qualités « stratégiques » enfin car les genres textuels actualisent parfaitement la toute dernière évolution de la communication interne et externe de l'entreprise dans une perspective intégrée de parties prenantes. Concept dynamique par excellence, le genre textuel fait office de pivot cognitif et communicationnel entre le contexte de l'entreprise, sa culture, sa politique et les textes produits par l'entreprise.

L'analyse des genres textuels privilégiés par l'entreprise a donc pour but d'élaborer un modèle de méta-données du genre nécessaires à la gestion des données lexicographiques et des documents qui constituent la base empirique du DE, et à élaborer les principes gouvernant la sélection des données lexicographiques elles-mêmes, extraites de ces mêmes genres. L'analyse des genres est enfin complétée par un audit lexicographique qui permet d'établir l'ordre prioritaire à attribuer aux fonctions lexicographiques du DE, à leur évolution et à leur remise à jour continue.

3. PRATIQUE DU « DE »

3.1. Pompes : accès aux discours contractuels et promotionnels

Voici, pour le projet d'un dictionnaire français-anglais de l'entreprise Grundfos, constructeur de pompes, un article prototypique démontrant quelques aspects de la démarche pratique, qui laisse une large part à la syntagmatique et à l'alignement des ressources traductiques, et permet à l'utilisateur d'accéder aux discours contractuels et promotionnels de l'entreprise :

pompe de chantier DW	contractor pump DW
<p>Les pompes DW peuvent transporter tous les solides pouvant passer par la crépine d'aspiration. La conception compacte de la pompe permet à la fois une installation temporaire et permanente. Grâce aux matériaux légers tels que les composites et l'aluminium, la pompe est facilement transportable.</p> <p>caractéristiques et avantages</p> <p>Très robuste grâce au choix de matériaux Maintenance simplifiée</p> <p>fonction et application</p> <p>Transfert de liquide dans les applications suivantes :</p> <p>Tunnels Mines Chantiers Carrières Étangs Sites de construction</p> <p>données techniques</p> <p>Débit, Q : maxi 300 m³/h Hmt, H : maxi 100 m CE Temp. du liquide : 0 °C à + 40 °C</p> <p>▲ Les pompes sont conçues pour un fonctionnement continu.</p> <p>→ Illustrations → Fiche technique → Manuel d'installation et de maintenance → Case story</p>	<p>The DW pumps can handle all solids that can pass through the suction strainer. The compact design makes the pumps suitable for both temporary and permanent installation. As the pumps are made of light materials such as composite and aluminium, they are easy to move from one place to another.</p> <p>features and benefits</p> <p>Extremely hard-wearing due to specially selected materials Service-friendly</p> <p>function and application</p> <p>The pump is suitable for liquid transfer in</p> <p>Tunnels Mines Quarries Gravel pits Fish ponds Building sites</p> <p>technical data</p> <p>Flow, Q : max. 300 m³/h Head, H : max. 100 m Liquid temp. : 0 °C to + 40 °C</p> <p>▲ Pumps are designed for continuous operation.</p> <p>→ Illustrations → Data sheet → Installation and maintenance instruction → Case story</p>

**Exemple 1 : Article prototypique du Dictionnaire Grundfos français-anglais
(source : Grundfos)**

Sélection lemmatique et lemmatisation : Pour faciliter l'accès, toutes les pompes de la gamme du constructeur sont sélectionnées et lemmatisées conformément à leur désignation et à leur classification au sein du groupe. La macrostructure du dictionnaire, du

moins en ce qui concerne les modèles de la gamme, reprend alors la forme du répertoire de catalogue.

Champs d'indication : Les champs d'indication de la microstructure incluent une définition à caractère fonctionnel, une rubrique avantages, une rubrique application, des données techniques, les collocations sous forme propositionnelle, et un champ de renvois sous forme de liens aux documents associés dans les deux langues ainsi qu'aux illustrations contenant également la nomenclature des pièces. Les champs d'indication regroupent des données permettant de desservir les fonctions traduction et connaissance du produit. Les données phraséologiques et syntagmatiques sélectionnées sont caractérisées par la double représentation d'unités provenant à la fois de discours contractuels (contenant les unités terminologiques classifiantes) et de discours promotionnels (contenant les arguments et propositions de vente etc.) selon la terminologie proposée par Cortès (2004) et reprise dans Leroyer (2005b). Toutes les données lexicographiques sont par ailleurs extraites des genres de la notice technique et de la rubrique catalogue, ainsi que du genre des « case stories » ou projets, qui sous forme narrative mettent en scène les scénarios de fourniture du modèle spécifique à des clients spécifiques. Au total, la structure de distribution des données lexicographiques démontre qu'il n'est pas nécessaire, comme le fait Wiegand (1994), d'opérer une distinction mouvante entre représentation de connaissances sémantiques et de connaissances encyclopédiques. Pour l'utilisateur du dictionnaire d'entreprise, seule la fonction lexicographique importe. La subdivision de l'article en champs fonctionnels facilite l'accès interne, et permet simultanément de réduire les coûts d'information lexicographique.

3.2. Éoliennes : Accès aux stratégies rhétoriques et aux objectifs de communication

Tous les genres textuels sont naturellement structurés en séquences de macrotraits linéaires, obligatoires et constitutifs du genre, qui facilitent grandement le découpage et le balisage des genres textuels en macrosegments textuels correspondants. Ces segments constituent la base de sélection des exemples textuels, dont les règles ont été formulées dans Leroyer (2002, 2005a) : il s'agit dès lors d'appliquer une grille d'analyse communicative des différents genres de l'entreprise, et de repérer les segments mettant en scène les stratégies rhétoriques privilégiées par l'entreprise. Dans un communiqué de presse mentionnant la défaillance d'une éolienne par exemple, où l'incident serait le résultat d'une maintenance déficiente, on isolera la section centrale du communiqué explicitant la stratégie de communication de crise qui consiste en la matière à nier la responsabilité du constructeur ; les balises < dénégation > permettront alors à l'utilisateur du DE de retrouver tous les contextes argumentatifs pertinents dans une situation de crise similaire, et donc de faciliter la production de texte et la formulation et transcription des messages appropriés (voir exemple 2 ci-dessous).

Le lundi 9 septembre 2002, une rupture de l'arbre principal sur un aérogénérateur de 500 kW doté d'un rotor de 43 mètres de diamètre a été rapportée sur une turbine installée en Allemagne depuis 1996. Cet incident n'a fait aucune victime et n'a causé aucun dommage corporel. < dénégation > Une enquête sur site a révélé que la rupture était clairement induite par une procédure de réparation déficiente causant un affai-

blissement du matériau de l'arbre principal et d'une qualité inadéquate de l'intervention elle-même... Cet incident est clairement isolé... < dénégation > [...]

Exemple 2 : Extrait de communiqué de presse (Source : Vestas)

Il va sans dire que la même démarche s'applique à tous les genres textuels privilégiés par l'entreprise, dans lesquels les stratégies rhétoriques peuvent être facilement identifiées et indexées au moyen de raffinements des métadonnées descriptives des genres textuels. Il en va de même pour l'indexation des objectifs de communication inscrits dans les messages génériques de l'entreprise – reconnaissance, notoriété spontanée, valeurs de la marque, justification des promesses, etc. – dont les étiquettes descriptives peuvent être fournies par les grilles d'analyse établies par Libaert (2005) pour la communication d'entreprise, et par Bhatia (2004) pour l'analyse des discours écrits.

4. PILOTAGE DU « DE »

Le DE n'est pas seulement un outil lexicographique polyfonctionnel, c'est aussi un outil stratégique pour l'entreprise. La conception et l'utilisation du DE doivent être basées sur un ensemble de décisions stratégiques cruciales gérant la prise en compte de toutes les parties prenantes. Préalablement à la prise des décisions lexicographiques, il est nécessaire de conduire une analyse poussée de la configuration des parties prenantes et des priorités lexicographiques à accorder : représentation des champs d'activités, et représentation des champs d'application des langues utilisées au sein de l'entreprise et par ses filiales internationales en particulier. Cette analyse doit inclure des enquêtes préconceptionnelles et des tests pilote au niveau des utilisateurs aussi bien qu'au niveau de la direction de l'entreprise, voir aussi Simonsen (2002b), qui discute le choix des méthodologies à mettre en œuvre aux différents stades d'un projet de DE. Dès lors que les parties sont impliquées dans le projet de DE et que leur expertise est exploitée au bon moment, le DE ne s'inscrit pas seulement dans l'organisation de l'entreprise, mais devient également un véritable outil de création de valeur pour les utilisateurs et pour l'entreprise.

Ce qui importe en lexicographie d'entreprise est de permettre aux utilisateurs de partager leurs connaissances et de commenter les formes (conception du système) et les contenus (données lexicographiques). Il s'agit donc bien d'établir une véritable communauté d'utilisateurs et d'éditeurs, voir à ce sujet Simonsen (2002a et 2002b) pour la définition des concepts de « participation des utilisateurs » et de « démocratie lexicographique » à tous les niveaux de conception et de gestion du projet de DE.

4.1. Un outil prescriptif et politique

Nous recommandons que les contenus du DE soient résolument prescriptifs, et qu'ils revêtent un aspect normalisateur en vue de supporter, faciliter et expédier la communication d'entreprise, en conformité avec la politique linguistique de l'entreprise et sa politique de communication, cf. à ce sujet Leroyer (2005a).

Suite à l'audit lexicographique, nous recommandons en outre que les règles s'appliquant à la mise en œuvre de la politique linguistique et de la politique de communication soient intégrées dans le DE. Dans la pratique, ceci revient à l'utilisation de référentiels de métadonnées du système de gestion des contenus telle que celle-ci est expliquée dans Simonsen (2005), ou bien, plus simplement, à la mise en place d'hyperliens spécifiques reliant les articles et composants internes et externes du dictionnaire aux éléments des plans et codes de l'entreprise correspondant. En ce sens, le DE se dote d'une dimension politique explicite.

4.2. Technologie du DE

La construction du DE impose le choix de solutions technologiques appropriées. La meilleure solution consiste à utiliser la plate-forme intranet de l'entreprise comme support du dictionnaire, puisque la plupart des entreprises sont dotées d'un intranet, et que tous les employés y ont accès et sont familiarisés à son utilisation quotidienne. De plus, l'intranet est protégé par un pare-feu protégeant les données sensibles. Nous recommandons également l'utilisation de pages ASP dynamiques associées à de puissantes bases de données relationnelles pour la saisie, l'extraction, la suppression et la remise à jour des données à partir des pages dynamiques. Nous préconisons finalement l'utilisation d'un système de gestion des contenus permettant à l'éditeur en chef et aux éditeurs délégués de gérer la configuration des droits d'accès des différents types d'utilisateurs dans l'optique d'une lexicographie véritablement participative selon Humbley (2002). Finalement, la gestion des contenus sera régie par un référentiel de métadonnées respectant les taxinomies de l'entreprise (hiérarchie, services, produits, etc.) et ses productions textuelles (genres et exemples textuels extraits de ces genres). L'architecture du module d'entrée du DE est illustrée dans la figure 1 ci-dessous.

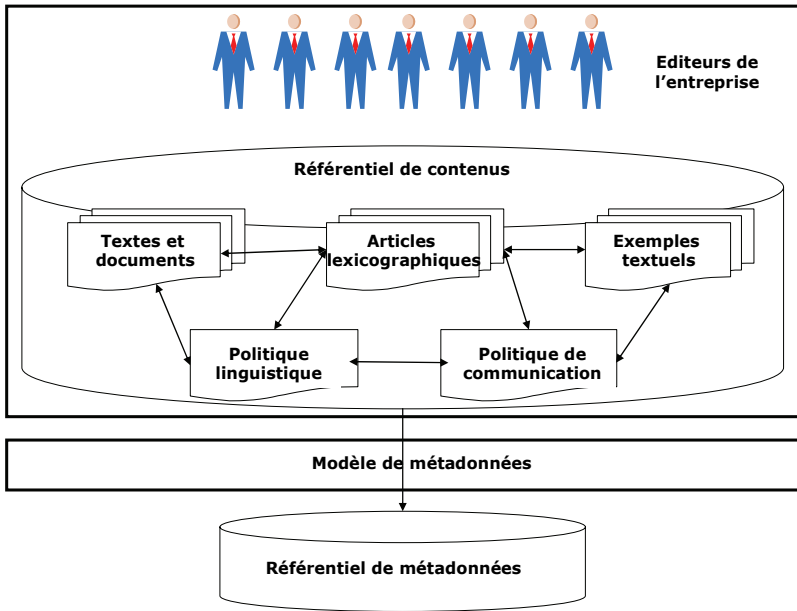


Figure 1 : Architecture du module d'entrée

Les documents créés étant encodés au format XML et automatiquement dotés de métadonnées descriptives, il suffit alors aux auteurs et éditeurs d'assigner à ces mêmes documents un jeu de métadonnées automatiquement proposé par le modèle du DE sous forme de menu déroulant. Les quelques minutes attribuées à cette tâche sont à terme génératrices de forts gains de temps pour l'éditeur en chef du DE.

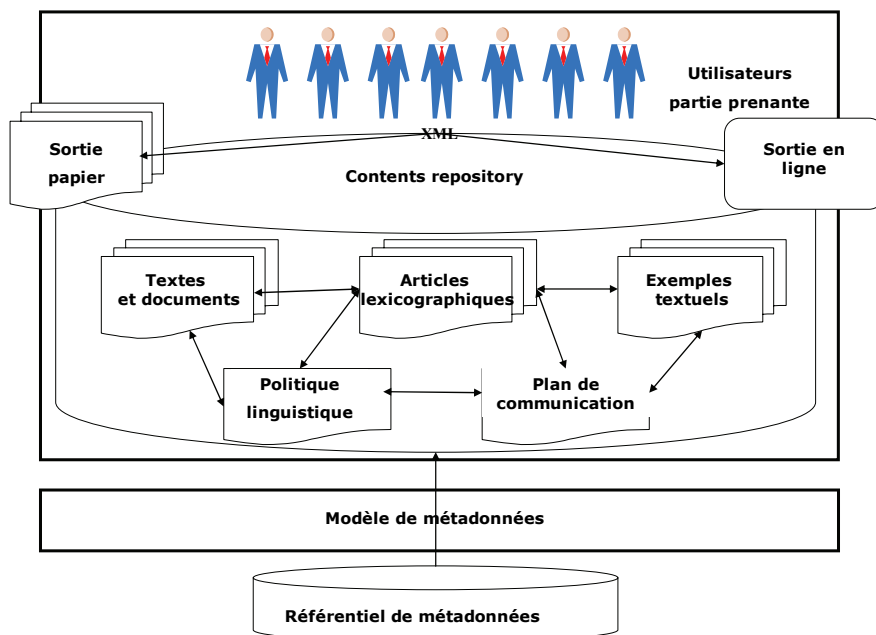


Figure 2 : Architecture générale du module de sortie

Le module de sortie (figure 2 ci-dessus) est conçu pour afficher les résultats de requête suivant les options et critères de recherche retenus par l'utilisateur, en ligne ou sur formulaire imprimé, grâce à la mise en relation structurée des différents modules par le biais des métadonnées du modèle et du référentiel.

L'exploitation optimale du système de gestion des contenus nécessite également la reprise de l'existant : les documents candidats non codés seront alors encodés et sauvegardés au format XML. Nous proposons une démarche semi-automatique permettant d'assigner à chaque document un jeu de méta-données fournies par le modèle.

4.3. Organisation du DE

Le DE doit être ancré au plus haut niveau de la direction de l'entreprise. Du point de vue organisationnel, nous proposons que l'éditeur en chef du DE soit placé dans le département de la communication et dépende directement du directeur de la communication. L'éditeur en chef désignera les experts ou éditeurs des différents départements, y compris les experts internationaux des filiales ou des partenaires de l'entreprise, à qui il appartient de valider les données encyclopédiques et linguistiques. La décision finale concernant la mise en relation des données lexicographiques aux documents, genres, et fragments de genres sous forme d'exemples textuels appartient à l'éditeur en chef. Cette philosophie de délégation lexicographique dans la création, la diffusion et le partage des connaissances est conforme aux principes établis par Nonaka (1994).

L'architecture modulaire présentée dans les figures 1 et 2 ci-dessus permet également de piloter la hiérarchisation des droits de rédaction, d'utilisation et de gestion du DE conformément à la politique organisationnelle de l'entreprise et à sa politique de sécurité de l'information. Dans l'ordre croissant, on aura ainsi : utilisateurs enregistrés, auteurs, éditeurs, publicateurs, managers, administrateurs et super-administrateurs.

4.4. Accès dynamique aux *corpus*, aux contextes et au web

Les principes d'accès dynamique aux *corpus* ont été développés dans Simonsen (2002a) pour la construction et l'installation du DE bilingue *TeleLex*, destiné aux employés de TDC, premier opérateur téléphonique danois. *TeleLex* permettait déjà aux utilisateurs du dictionnaire de consulter directement et dans les deux langues les *corpus* alignés sous-jacents (DANCORP et USCORP) constituant la base empirique du dictionnaire, tout en démarrant automatiquement un concordancier à partir de l'article du dictionnaire.

Ces principes ont été repris et raffinés dans la conception et dans la compilation de *ZooLex*, un DE multilingue destiné aux utilisateurs du jardin zoologique de Copenhague (voir l'appendice 1, qui présente des captures d'écran illustrant les différentes options d'accès aux contextes). Dans la pratique, les utilisateurs consultant les articles de *ZooLex* disposent d'un accès structuré à l'information contextuelle stockée et indexée sur le serveur de l'intranet du zoo – dans les lettres d'information par exemple pour la communication marketing – de même qu'ils peuvent à tout moment consulter directement, à partir des articles du dictionnaire, les grands *corpus* de langue générale danois (*Korpus 2000*) et britannique (*British National Corpus*), ou bien lancer une recherche sur Google en double cliquant sur les termes ou les collocations.

4.5. Métadonnées et interface

Nous proposons que l'utilisateur dispose d'options de recherche avancées pour accéder aux documents et aux exemples textuels indexés sur la base des genres textuels, du récepteur, du type de projet, etc. Ceci nécessite le recours à un modèle de métadonnées qui seront assignées manuellement, semi-automatiquement ou automatiquement aux documents en conformité avec la taxinomie de l'entreprise.

Le modèle (voir appendice II en anglais) est dérivé du « Dublin Core » et comprend deux catégories de métadonnées, avec leurs éléments et leurs raffinements au niveau des genres textuels : métadonnées cycle de vie et méta-données classifiantes permettant d'identifier efficacement les différents contextes dynamiques. On y retrouve en effet toutes les catégories mises en place par « l'initiative Dublin Core » (titre, créateur, sujet, description, publicateur, contributeur, date) et de nouveaux raffinements (type de texte, situation de communication). Une fois mis en place, le recours aux métadonnées permet à l'utilisateur de lancer au moyen d'un simple menu déroulant des recherches précises, et de demander au moteur du DE de retrouver tels genres de textes, tels auteurs, tels contenus reliés à tel projet, ou bien telles exemplifications de stratégies spécifiques de communication *corporate* grâce au découpage et au balisage des genres textuels en macrosegments comme il a été démontré plus haut.

Conformément à la philosophie fonctionnelle du DE, l'interface de requête doit finalement proposer à ses utilisateurs des options de recherche et d'affichage à géométrie variable et individualisées, avec une disposition dynamique des pages écran facilitant l'accès selon les besoins.

5. CONCLUSION

« La plus grande force d'un dictionnaire se situe au niveau du mot. Au niveau de la collocation et de l'idiome, un dictionnaire peut rester un outil utile. Au niveau de la phrase, son utilité est déjà très limitée, mais au niveau textuel il n'a plus aucune utilité. »

On a longtemps estimé, comme Tarp (1999) cité ici, que les dictionnaires étaient logiquement basés sur l'inventaire et la description systématique et exhaustive des unités lexicales ou terminologiques. Et que les dictionnaires étaient incapables d'opérer au niveau textuel. En ce sens, les contextes venaient enrichir et affiner les descriptions adaptées aux besoins des utilisateurs mais ne constituaient pas, faute d'accès, des données lexicographiques et des champs d'indication en tant que tels.

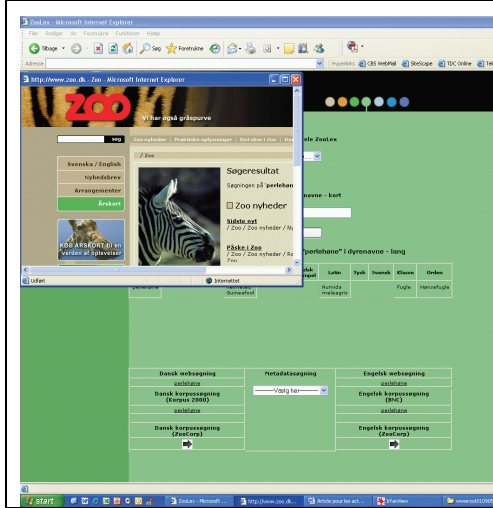
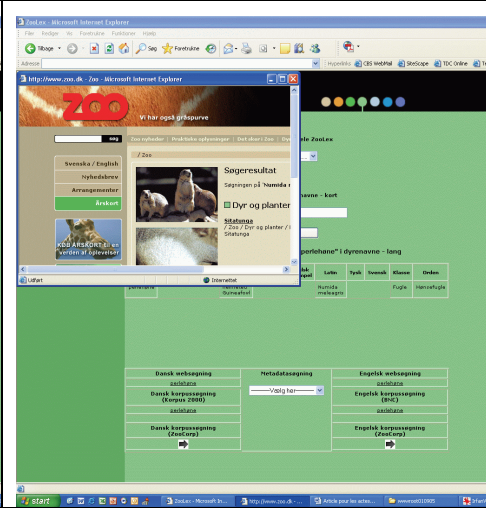
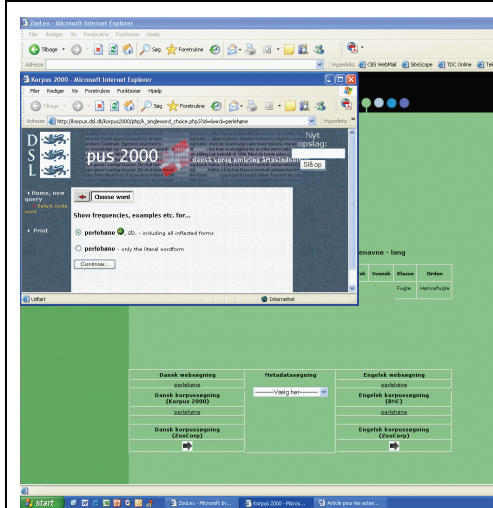
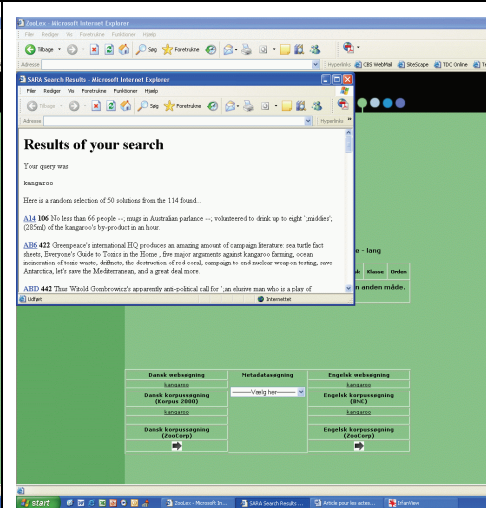
La théorie et la pratique du DE que nous avons brièvement introduites ici illustrent la démarche inverse. Les fonctions lexicographiques sont dérivées des besoins textuels des utilisateurs dans le contexte de communication de l'entreprise. Ce sont les textes, dont la création et l'utilisation sont régies par la reconnaissance et par la prédictibilité énonciative des genres textuels, qui doivent encadrer toute entreprise lexicographique ou terminologique au sein de l'entreprise. Les utilisateurs en effet n'ont pas seulement besoin des termes et des relations conceptuelles qui en modélisent le sens ; ils ont aussi besoin des textes et des contextes qui correspondent parfaitement à l'usage et à l'actualisation de ces termes. Les traducteurs en particulier, c'est un truisme de le dire, ne traduisent pas des termes, mais des textes. Dans le DE, l'intégration dynamique des genres textuels comme base empirique et structurante du dictionnaire permet de refléter la réalité opérationnelle et prescriptive de l'entreprise, conformément à son contexte d'utilisation et de communication et de mieux répondre fonctionnellement parlant aux besoins spécifiques des utilisateurs en exploitant les potentiels des dernières technologies de l'information et de la communication. Dans son ontogénèse, le DE a ainsi pour vocation et mission de consacrer le mariage des sciences du langage, de la communication d'entreprise et de la lexicographie fonctionnelle au profit de ses utilisateurs.

BIBLIOGRAPHIE

- Bergenholtz (H.), Tarp (S.), 2003 : « Two opposing theories : On H.E. Wiegand's Recent Discovery of Lexicographic Functions », dans *Hermes Journal of Linguistics*, 31, pp. 171-196.
- Bhatia (V. K.), 2004 : *Worlds of Written Discourse*, London and New York, Continuum.
- Cortès (M.), 2004 : « Terminologie et syntaxe de la classifiante », dans Cortès (C.), édit., dans *Des Fondements théoriques de la terminologie*, Cahier du CIEL. 2004, Paris, Université de Paris VII, Denis Diderot, UFR EILA.

- Humbley (J.), 2002 : « Nouveaux dictionnaires, nouveaux rapports avec les utilisateurs », dans *Meta*, XLVII, 1, pp. 95-104.
- Lerat (P.), 1995 : *Les langues spécialisées*, Paris, Presses Universitaires de France.
- Leroyer (P.), 2002 : *Les Exemples textuels unificateurs. Fondements et Applications pour le Dictionnaire de Traduction Technique*, Århus, The Århus School of Business.
- Leroyer (P.), 2003 : « Dictionnaire d'entreprise : L'exemple des éoliennes », dans Szende (T.), édit., *Actes des journées de lexicographie bilingue*, INALCO, Paris, H. Champion, pp. 395-407.
- Leroyer (P.), 2005a : « Bringing corporate dictionary design into accord with corporate image : From words to messages and back again », dans Gotlieb (H.) et Mogensen (J. E.), édit., *Dictionaries, Lexicographical options and User Needs*, John Benjamins Publishing Company, Terminology and Lexicography Research and Practice, Amsterdam, (à paraître).
- Leroyer (P.), 2005b : « Terminologi, se leksikografi, leksikografi se ordbog », [compte rendu de] Cortès (C.), éd., *Des fondements théoriques de la terminologie*, Cahiers du Ciel 2004, Paris, Université de Paris VII, Denis Diderot, UFR Eila, dans *LexicoNordica* 12, (à paraître).
- Libaert (T.), 2003 : *Le Plan de communication. Définir et Organiser votre stratégie de communication*, Paris, Dunod.
- Nonaka (I.), 1994 : « A Dynamic Theory of Organizational Knowledge Creation », dans *Organization Science*/vol. V, n° 1., pp. 14-37.
- Simonsen, (H. K.), 2002a : « *TeleLex – Theoretical Considerations on Corporate LSP Intranet Lexicography : Design and Development of TeleLex – an Intranet-based Lexicographic Knowledge and Communications Management System* », Århus, École des hautes études commerciales d'Århus.
- Simonsen, (H. K.), 2002b : « User Involvement In Corporate LSP Intranet Lexicography », dans Gotlieb (H.), Mogensen (J. E.) et Arne Zettersten (A.), édit., *Proceedings of The eleventh international Symposium on Lexicography*, May 2-4, 2002 at the University of Copenhagen, Lexicographica, Series Maior 115, pp. 489-509.
- Simonsen, (H. K.), 2004 : « Nine Key Principles on Corporate LSP Intranet Lexicography », dans *Proceedings of the Eleventh Euralex international Congress, 6-10 July 2004*, Université de Bretagne Sud, Lorient, vol. 2, pp. 603-613.
- Simonsen, (H. K.), 2005 : « Corporate Communication and Corporate Lexicography : A Value-Adding Combination », dans *Proceedings of 7th European Convention – Business Communication : Making an Impact, May 2005*, Copenhagen Business School (à paraître).
- Tarp (S.), 2001 : « Lexicography and the linguistic concepts of homonymy and polysemy », dans *Lexicographica* 17 : 22-39.
- Wiegand (H. E.), 1994 : « Zur Unterscheidung von semantischen und enzyklopädischen Daten in Fachwörterbüchern », dans Schaefer (B.) et Bergenholtz (H.), édit., *Fachlexikographie, Fachwissen und seine Repräsentation in Wörterbüchern*, Tübingen, Gunter Narr Verlag, pp. 103-159.

Appendice 1 – Captures d’écran du dictionnaire *ZooLex*

<p><i>ZooLex</i> – popup document search (lexical item)</p>	<p><i>ZooLex</i> – popup document search (multiword lemma)</p>
 <p>The screenshot shows the ZooLex website interface. A search bar at the top contains the word 'Zoo'. A popup window titled 'Søgesultat' (Search result) is displayed, showing the search term 'Zoo' and its definition: 'Zoo / Zoo dyr eller / Plakater'. The website has a green background with various navigation and search options.</p>	 <p>This screenshot is similar to the previous one but shows a different search result. The popup window displays 'Søgesultat' for 'Zoo' and lists related terms like 'Dyr og planter' and 'Matstua'. The website layout and search bar are consistent with the previous image.</p>
<p><i>ZooLex</i> – Danish Corpus Search – <i>Korpus 2000</i></p>	<p><i>ZooLex</i> – English Corpus Search – <i>BNC</i></p>
 <p>The screenshot shows the 'Korpus 2000' search interface. A search bar contains the word 'Zoo'. The results show a list of words and their frequencies, including 'Zoo' and 'Dyr'. The interface is designed for searching within a large Danish corpus.</p>	 <p>This screenshot shows the 'BNC Search Results' interface. The search bar contains 'Zoo'. The results section is titled 'Results of your search' and shows a list of words and their frequencies, including 'Zoo' and 'Dyr'. The interface is designed for searching within the British National Corpus.</p>

ZooLex – metadata search

The screenshot shows the ZooLex metadata search interface. The browser window title is "ZooLex - Microsoft Internet Explorer". The address bar shows "http://www.zoolex.dk". The page content includes a navigation menu on the left with links like "Home - ZooLex", "Dyrescience - kort", "Dyrescience - lang", "Drebnag - kort", "Drebnag - lang", and "ZooLex-differ". The main content area has a search bar with the text "Vejrg huse" and a dropdown menu for "Vejrg huse" with options: "Fæl", "Fælde", "Fældeord", "Dialektisk", "Oversættelse af", "Multiord", "Gabe", "Bredde", "Form", "Inocula", "Spring", "Hænder", "Håning for", "Puffgæder", "Egnskab", "Hænder", "Lodmodning", and "Tauske". Below the search bar are buttons for "Dansk webopslagning", "Dansk korpusopslagning (Egenskab)", "Dansk korpusopslagning (Drebnag)", "Engelsk webopslagning", "Engelsk korpusopslagning (Hæ)", and "Engelsk korpusopslagning (Egenskab)".

Appendice 2 – Mod le de m tadonn es du DE

<p>Page title : (Automatic entry of default value describing the title of the resource)</p>	<input type="text"/>
<p>Alternative title : (Manual entry of other titles than the main title)</p>	<input type="text"/>
<p>Creator : (Automatic entry of default value describing the name of the author or company responsible for the creation of the content – last name – first name – e-mail)</p>	<input type="text"/>
<p>Subject : (Manual entry of subject keywords describing the content)</p>	<input type="text"/>
<p>Subject : (Semi-automatic selection of taxonomy keywords from pre-defined schemes)</p>	<input type="text" value="-----Select keyw ords-----"/>
<p>Description : (Manual entry of abstract or content description of ressource)</p>	<input type="text"/>
<p>Publisher : (Automatic entry of default value describing the name of the business unit, department, division, or function)</p>	<input type="text"/>
<p>Contributor : (Manual entry of people or organizations, who have contributed to the resource in question)</p>	<input type="text"/>
<p>Date : (Automatic entry of default 8-digit number describing the date on which the resource was published)</p>	<input type="text"/>
<p>Text Type : (Semi-automatic entry of the category or genre of the resource in question, for</p>	<input type="text" value="-----Select text type-----"/>

example technical report, press release, marketing material etc.)	
Communication situation : (Semi-automatic entry of the communication situation where the resource is to be used)	-----Select text type-----
Format : (Automatic entry of default value describing the format of resource, for example.htm, .html, .doc, .pdf, etc.)	
Life-cycle : (Manual entry of lifecycle data describing what to do with the resource when the cutoff data is exceeded)	
Language : (Semi-automatic entry of the language of the content)	-----Select language-----
Relation : (Manual entry of relation to other resources)	
Coverage : (Manual entry of the spatial, organizational, and temporal characteristics for the resource in question)	
Rights : (Automatic entry of default value describing the copyright rules of the resource)	
Ownership : (Manual entry of ownership data on read/write rights)	
Version : (Semi-automatic entry of version of resource)	-----Select version type-----
Approval : (Semi-automatic entry of approval workflow – who is responsible for the approval and publication)	-----Select workflow -----

L'EMPLOI DE TERMES RÉDUITS COMME RÉVÉLATEUR DE LA CENTRALITÉ DANS LE DOMAINE

Marie-Paule Jacques

ERSS – Université Toulouse II-le Mirail
Toulouse, France.

Résumé : *On s'intéresse ici aux formes identiques à la tête d'un terme complexe qui peuvent être considérées comme un terme réduit, c'est-à-dire l'équivalent d'un terme complexe mais amputé de son expansion. En chiffrant les occurrences des deux types d'emploi de ces termes réduits, soit comme reprise anaphorique, soit sans antécédent textuel, on constate que certains se dispensent très volontiers de leur terme complexe source, tout en se présentant avec un déterminant défini. Nous expliquons ceci par la notion de centralité.*

Mots-clés : *réduction de termes complexes, analyse en discours, centralité*

1. INTRODUCTION

Ce n'est que récemment que les linguistes et terminologues se sont avisés de replacer les termes dans leur contexte et de s'intéresser à leurs usages en situation réelle de communication spécialisée. Ce faisant, il apparaît que les termes ne sont pas que des étiquettes pour des concepts, ce sont aussi des éléments des textes et des discours, donc des formations langagières plus malléables qu'on ne voulait le reconnaître d'abord. Et en tant que tels, les termes sont modifiés, raccourcis, tronqués, simplifiés, réduits. Précisément parce qu'ils sont immergés dans un co-texte, dans un contexte qui permet et la réduction des termes, et l'interprétation de leurs formes réduites, cf. Collet (2000), Freixa (2002).

Nous présentons ici quelques-uns des résultats obtenus en analysant les phénomènes de réduction des termes complexes dans deux *corpus* de textes spécialisés. Le propos général de l'étude était de montrer comment l'immersion dans un discours permet d'omettre une partie *a priori* essentielle d'un terme complexe – c'est-à-dire formé de plusieurs mots, comme par exemple *contrôleur de carrefour*, *équipement de terrain*, *effort à la commande*, *réseau routier national*, etc. Selon les contextes, de tels termes peuvent se voir réduits à leur tête ou à leur expansion (signalons que la réduction ne se limite pas à la suppression des constituants majeurs du terme complexe, mais nous choisissons de nous intéresser seulement à ces effacements). Dans cet article, nous nous focaliserons sur le premier type de réduction, qui ne laisse à la surface du texte que la tête d'un terme complexe, soit *contrôleur*, *équipement*, *effort*, *réseau*, pour les exemples cités plus haut. Nous nous attacherons à mettre en évidence des différences de comportement des termes complexes à l'égard de cette réduction, différences perceptibles sur le plan quantitatif et qui révèlent selon nous des statuts différents des no-

tions dénommées au sein du domaine. La proposition que nous illustrerons est l'idée que, plus les formes réduites considérées sont autonomes dans les textes, plus les notions dénommées par les termes complexes correspondants sont centrales dans le domaine.

Il va de soi qu'une telle étude n'est possible que par l'analyse de *corpus* textuels. Nous présenterons dans la section 2 ces *corpus* et la méthode d'analyse. La section 3 sera consacrée aux résultats chiffrés et à la défense de notre hypothèse, la section 4 proposera, en guise de conclusion, une réflexion sur ce que peut apporter une analyse textuelle du fonctionnement des termes.

2. CORPUS ET MÉTHODE

Notre objet d'étude réside principalement dans l'analyse des phénomènes de réduction de termes complexes dans leur usage réel, c'est-à-dire dans les textes produits par et pour les experts de domaines spécialisés, dont le propos fait donc crucialement intervenir les notions et les dénominations propres aux domaines concernés. Afin de ne pas nous limiter à un seul domaine, nous avons mené notre étude sur deux *corpus*. Leur caractéristique commune est de servir une communication entre experts, mais ils proviennent de domaines différents et de canaux de publication différents : le *corpus* professionnel appartient au domaine de la gestion des déplacements et est constitué de textes professionnels à diffusion restreinte ; le *corpus* sport-loisir appartient au domaine du vol libre et est constitué d'articles de magazines spécialisés à diffusion non restreinte : tout un chacun peut se les procurer dans n'importe quelle librairie.

Ces deux *corpus* ont été analysés avec *Syntex* (Bourigault *et al.* 2005), qui produit, entre autres résultats, une liste de candidats termes simples et complexes. Parmi ceux-ci, nous avons choisi une centaine de termes complexes dans le *corpus* professionnel et une quarantaine dans le *corpus* sport-loisir, en privilégiant deux critères pour le choix des termes à analyser : la fréquence et le partage de la tête avec un autre terme. En effet, pour la compréhension des mécanismes de réduction, il s'avère plus pertinent d'analyser les situations dans lesquelles une certaine forme est susceptible de renvoyer à plus d'un terme complexe, car ces situations permettent de cerner pourquoi et comment le locuteur d'un domaine spécialisé, en un certain point de son discours, peut se dispenser de la part d'information contenue dans l'élément effacé du terme complexe. Par exemple, dans le *corpus* professionnel, un seul terme, *contrôleur de carrefour*, comporte comme tête syntaxique le mot *contrôleur*. De ce fait, l'emploi de ce mot seul renvoie inmanquablement au terme *contrôleur de carrefour* :

Au PC Capitoul, la régulation en mode automatique mise en suvre par Capitoul consiste, zone par zone, à : [...] envoyer à *chaque contrôleur* concerné les commandes correspondant aux plans de feux [...]

Comme il n'y a aucune autre possibilité, le scripteur n'a à s'entourer d'aucune précaution particulière, l'interprétation est en quelque sorte garantie par l'unicité du terme source potentiel. Mais pour le linguiste ou le terminologue soucieux de saisir le fonc-

tionnement langagier des termes complexes et leurs modalités de réduction, une telle situation n'est guère informative.

Plus riche d'informations est en revanche le cas de figure dans lequel une forme simple correspond à la tête de plusieurs termes complexes, comme – toujours dans le *corpus* professionnel – le mot *dossier* qui peut renvoyer à *dossier [d'avant-projet sommaire]*, *dossier [de consultation des entreprises]*, *dossier [de voirie d'agglomération]*, *dossier [d'études préliminaires]* ou encore *dossier [d'exploitation]*. En ce cas, lorsque le mot *dossier* apparaît sans aucune expansion (l'expression entre crochets), le discours doit comporter les moyens d'une interprétation correcte. Celle-ci est à coup sûr obtenue si la forme simple fait suite à une mention du terme complexe, c'est-à-dire si elle est anaphorique :

Le dossier d'exploitation est transmis [...] au plus tard trois semaines avant la date prévue de début du chantier. Ce dossier est établi en deux exemplaires, dont un est conservé par le gestionnaire de la voirie.

Mais, et cette situation est la plus intrigante, il peut arriver que la forme simple soit employée hors reprise anaphorique, non pour évoquer la notion hyperonyme, mais bien avec la valeur définie par l'un des termes complexes du domaine. On trouvera par exemple :

Les données de trafic sont recueillies en temps réel au moyen de stations de comptages régulièrement réparties sur l'ensemble *du réseau*.

où *réseau* doit être compris comme *réseau de l'agglomération* (toulousaine), sans cependant que ce terme apparaisse antérieurement.

Afin de déterminer comment le co-texte restitue cette valeur, afin de saisir les différents mécanismes à l'œuvre, nous avons analysé manuellement toutes les occurrences de formes simples identiques à la tête d'une série de termes complexes (par exemple *bulletin, chantier, comité, dossier, équipe, équipement, réseau, système...* pour le *corpus* professionnel ; *aile, appui, ascendance, bord, effort, fermeture...* pour le *corpus* sport-loisir).

Sans détailler ici tous les aspects de l'analyse, il faut préciser que chaque occurrence a donné lieu à un classement : il faut d'abord déterminer si elle a ou non la valeur d'un terme complexe – ce que nous apprécions principalement par la possibilité de substituer le terme complexe à l'occurrence, cf. Jacques (2003a), Jacques (2003b) –, si oui, nous la considérons comme un *terme réduit* et nous la relierons à son terme complexe source ; ensuite déterminer si le terme complexe source est présent dans l'environnement du terme réduit ou non. Si oui, nous parlons de *reprise anaphorique* (RA), si non, de *termes réduits sans antécédents textuels* (TSA). Nous nous intéressons dans la section qui suit à quelques caractéristiques remarquables de ces deux catégories, en présentant une partie des résultats de l'analyse.

3. RÉSULTATS DE L'ANALYSE

Commençons par des résultats chiffrés. Le tableau qui suit indique, pour chaque *corpus*, le nombre d'occurrences de reprises anaphoriques (RA), le nombre d'occurrences de termes réduits sans antécédents (TSA) et le total de ces deux catégories, c'est-à-dire le nombre total d'occurrences de termes réduits analysées. Dans la dernière colonne, nous mentionnons à titre indicatif le nombre d'occurrences des termes complexes correspondants.

	RA	TSA	Total	Termes complexes
<i>Corpus professionnel</i>	313	675	988 (35 %)	1 822 (65 %)
<i>Corpus sport-loisir</i>	56	185	241 (28 %)	615 (72 %)
Total	369	860	1 229 (34 %)	2 437 (66 %)

Tableau 3 : Nombre d'occurrences analysées

Ce tableau fait apparaître deux faits d'importance. Premièrement, la réduction des termes n'est pas un phénomène marginal dans les textes, il ne s'agit pas ici de quelques occurrences isolées, mais bien d'un mécanisme systématique. Deuxièmement, la reprise anaphorique est finalement beaucoup moins présente que la réduction en l'absence du terme complexe source : seulement 30 % des occurrences de termes réduits sont des RA. Et l'on verra plus loin que cette disproportion globale s'accroît encore pour certains termes, et ce dans les deux *corpus*. Mais avant d'entrer dans ces considérations, analysons les indices par lesquels ces différents termes réduits peuvent être interprétés convenablement, c'est-à-dire les indices par lesquels un expert du domaine est capable de relier une forme simple à un terme complexe source.

3.1. Des indices pour l'interprétation des formes réduites

Lorsque le terme réduit suit de près le terme complexe source, son interprétation repose sur une mise en relation avec son antécédent. Une imposante littérature linguistique sur l'anaphore et la référence – Apothéloz (1995a), Apothéloz (1995b), Charolles (2002), Corblin (1987), Corblin (1995), Cornish (2001), Cornish (2003), de Mulder (1994), Kleiber (1986), Reichler-Béguelin (1988) – met le déterminant de l'expression anaphorique au premier rang des moyens par lesquels cette mise en relation est assurée. Effectivement, dans les deux *corpus*, la reprise anaphorique apparaît massivement avec un déterminant défini ou un démonstratif (78 % des RA dans le *corpus* professionnel, 56 % dans le *corpus* sport-loisir) qui signalent une certaine accessibilité du terme complexe. Le terme réduit, bien qu'incomplet, récupère de cette façon la valeur sémantique et référentielle du terme complexe antécédent.

Lorsque le terme réduit est employé sans mention antérieure du terme complexe, une telle opération de récupération *via* un antécédent est impossible. Le co-texte et le contexte de l'occurrence du terme réduit jouent alors un rôle majeur : ils délimitent un cadre dans lequel une certaine interprétation de la forme est plausible. La mise en place de ce cadre interprétatif repose tout à la fois sur des indices locaux tels que :

- la structure verbale dans laquelle le terme réduit est enchâssé ;
- le lexique environnant ;
- le thème du paragraphe ou de la section,

et sur des indices plus globaux tels que :

- la thématique globale du document ;
- la centralité du référent du terme dans le domaine.

La centralité du référent renvoie au fait que, dans une sphère d'activités donnée, certains référents sont plus au centre des préoccupations des locuteurs de cette sphère d'activité que d'autres référents (par exemple, *ascendance thermique* est une notion centrale dans le domaine du vol libre). Nous montrerons dans la prochaine section que nous la déduisons des types d'emplois des termes réduits dans les textes et que nous en faisons le motif d'une interprétation par défaut de la forme réduite, interprétation par défaut qui peut être annulée par les indices présents à la surface du texte.

La structure que le terme réduit forme avec le verbe de la phrase renseigne le lecteur sur sa référence.

[1] *L'inversion* demande 3 secondes. [*corpus sport-loisir*]

Dans cette phrase, *l'inversion* a la valeur du terme *inversion de virage*, ce qui apparaît nettement dès lors qu'on la rapproche de l'extrait suivant :

[2] *L'inversion de virage* demande 2,5 s à 30 km/h et 30° d'inclinaison. [*corpus sport-loisir*]

De manière tout aussi évidente pour les locuteurs du domaine, le lexique environnant l'occurrence détermine sa valeur sans ambiguïté. Dans l'exemple suivant, c'est l'autre valeur de *l'inversion*, *l'inversion de température*, qui est visée, ce qu'un expert du domaine saisit à la lecture de la combinaison du terme réduit avec *l'altitude*.

[3] En conclusion, cette journée du 10 août a été exceptionnelle en montagne grâce à la conjonction de conditions anticycloniques pas trop « bloquantes » et d'une forte chaleur en basses couches, permettant aux ascendances de dépasser largement l'altitude de *l'inversion*. [*corpus sport-loisir*]

Au-delà du lexique, la thématique du co-texte joue aussi un rôle primordial. Par exemple, dans l'extrait suivant, le titre recèle un indice pour l'interprétation d'un dossier :

[4] Rédaction de l'APS relative aux équipements terrain [titre]

Pour ce qui concerne les équipements terrain, leur identification est déjà pratiquement connue. Le travail à accomplir est donc essentiellement de monter un dossier définissant les spécifications techniques minimums, des schémas d'implantation et une approche financière. [*corpus professionnel*]

Sachant que le thème de cette section est la rédaction d'un avant-projet sommaire (APS), le *dossier* en question est un *dossier d'avant-projet sommaire* et non un *dossier d'exploitation* ou un *dossier de consultation des entreprises*.

À un autre niveau, plus éloigné de l'occurrence elle-même du terme, le thème global du document est un facteur d'interprétation primordial. Par exemple, dans les documents du *corpus* professionnel consacrés à la mise en place d'un système dit *système Erato* pour la gestion d'une partie du réseau routier appelée *réseau Erato*, on trouve 111 occurrences de *réseau* sans antécédent textuel. Sans surprise, 86 de ces occurrences sont la réduction du terme *réseau Erato*.

Le thème global du document fait de l'un des termes un élément central, ce qui est marqué par l'abondance des références à ce terme au moyen de descriptions sous-spécifiées, c'est-à-dire des occurrences de la tête du terme complexe sans antécédent textuel auquel se « raccrocher ». Nous pensons que, dans ce cas, la forme est traitée au moyen d'une interprétation par défaut qui serait « étant donné l'univers de discours construit par le texte, la forme *le N* renvoie, par défaut, au terme complexe *N-Y* », c'est-à-dire, par rapport à l'exemple utilisé ci-dessus : « étant donné que le document concerne au premier chef la mise en place du *système Erato*, la forme *le réseau* renvoie, par défaut, au terme complexe *le réseau Erato* », et ce, parce que *le réseau Erato* est, dans ces documents, un élément central :

« Il faut préciser que ce statut [de centralité] présente lui aussi, mais à un autre niveau, un caractère contingent : un objet n'est bien sûr pas central « en soi » ; il ne l'est que dans le cadre d'un texte donné, ou d'un secteur de ce texte. La propriété de centralité est liée à un projet de signification. » (Apothéloz 1995b : 316).

Il n'est pas anodin de constater que, malgré l'absence du terme complexe source, la majorité de ces occurrences est déterminée par un article défini (72 % pour le *corpus* professionnel, 52 % dans le *corpus* sport-loisir). Celui-ci constitue une indication au lecteur que la valeur de l'occurrence est aisée à récupérer, autant que pour une reprise anaphorique, et ce parce que le référent est aussi saillant que s'il venait d'être évoqué par un terme complexe explicite. Ceci s'explique si l'on considère qu'une sphère d'activité donnée privilégie certaines de ses notions spécifiques.

3.2. Les termes réduits sans antécédent textuel : un révélateur de la centralité dans le domaine

Nous proposons d'étendre ces deux notions de centralité du référent et d'interprétation par défaut au domaine, plus précisément à la sphère d'activité elle-même. Notre analyse met en lumière que, pour certains des termes complexes de chacun des deux domaines, il existe dans les textes un usage prépondérant du terme réduit soit comme reprise anaphorique, soit sans antécédent textuel. Le tableau suivant ne fournit pas une liste exhaustive de nos décomptes, celle-ci est présente dans Jacques (2003a), mais donne, pour certains termes, le nombre d'occurrences du terme complexe, de la reprise anaphorique et du terme réduit sans antécédent textuel correspondant.

	terme complexe	RA	TSA
Corpus professionnel			
caméras de terrain	2	0	16
chantier courant	7	10	0
chantier non courant	5	5	0
contrôleur de carrefour	84	33	15
équipement de terrain	55	20	82
exploitation de la route	38	12	81
réseau (urbain) de surface	16	0	1
réseau de l'agglomération	1	0	7
réseau Erato	28	5	89
réseau informatique	3	0	7
réseau routier	40	9	114
système Capitoul-2	28	5	64
système Erato	88	23	31
Corpus sport-loisir			
ascendance dynamique	1	0	0
ascendance thermique	14	44	77
effort à la commande	24	4	50
effort en virage	16	0	0
fermeture asymétrique	9	1	18
fermeture frontale	4	0	0
inversion de température	3	3	6
inversion de virage	26	0	5

Tableau 4 : Échantillon de la répartition de termes réduits entre reprises anaphoriques et termes réduits sans antécédent textuel

Par exemple, dans le *corpus* professionnel, la forme réduite de *caméras de terrain* (donc *caméras*), n'est jamais employée dans un co-texte de reprise anaphorique, mais toujours dans un co-texte où le terme n'est pas exprimé. À l'inverse, *chantier (s)*, forme réduite de *chantier courant* ou de *chantier non courant*, n'apparaît que dans des cotextes de reprise anaphorique. Pour les termes *réseau routier*, *réseau Erato*, *équipement de terrain*, *exploitation de la route*, *système Capitoul-2* ou encore *système Erato*, les deux types d'occurrences, comme reprise anaphorique ou sans antécédent textuel,

peuvent être observées dans les textes, mais avec une supériorité quantitative très nette des secondes.

De même, dans le *corpus* sport-loisir, le terme réduit *ascendance* comme réduction du terme complexe *ascendance thermique* est utilisé préférentiellement sans antécédent textuel. La notion dénommée *fermeture asymétrique* est plus volontiers désignée par le terme réduit sans antécédent textuel qu'en situation de reprise anaphorique, cependant que la forme *fermeture* n'est jamais employée pour un autre terme du domaine, *fermeture frontale*. Dit autrement, lorsque le scripteur des textes analysés parle de *fermeture asymétrique*, il emploie soit le terme complexe, soit *fermeture*, ce dernier préférentiellement à la place du terme complexe, rarement pour une anaphore. Mais lorsqu'il évoque la *fermeture frontale*, alors il emploie toujours le terme complexe, jamais une forme réduite. On pourrait faire le même type de remarque pour *effort à la commande* et *effort en virage* : le premier peut être simplement évoqué par le terme *l'effort* alors que le second est toujours mentionné avec tous ses constituants.

Nous voyons cette différence d'emploi comme un indice de la position du référent du terme dans le domaine : plus l'entité dénommée par le terme est centrale, c'est-à-dire fait l'objet des préoccupations des locuteurs du domaine, plus ceux-ci auront tendance à réduire le terme, quel que soit son contexte d'emploi. Et s'ils se permettent d'éliminer les éléments de spécification du terme complexe, c'est parce que la forme résultante sera interprétée par défaut comme correspondant à ce terme, comme si elle était univoque, comme, par exemple, *contrôleur* qui ne renvoie, dans le *corpus* professionnel, qu'à un terme complexe du domaine, *contrôleur de carrefour*. Dans la sphère d'activité de ces locuteurs, évoquer par exemple *le réseau* sous-entend qu'une des sortes de réseau possibles s'impose par défaut comme *le réseau*, et ce d'autant plus lorsque sont présents dans le co-texte des indices qui confortent cette interprétation par défaut, tels ceux que nous avons précédemment mis en lumière. Par exemple, il apparaît que *le réseau* sans expansion et sans antécédent textuel est essentiellement utilisé pour une référence au *réseau routier* (114 occurrences).

De même, dans le domaine du vol libre, parler de *l'ascendance*, c'est signifier que l'une des notions qui peut être désignée par ce syntagme nominal, *l'ascendance thermique*, est centrale dans le domaine. Quand le référent désigné occupe dans le domaine une place plus centrale qu'un autre référent qui pourrait être évoqué par le même terme réduit, il devient presque redondant de réaliser le constituant qui en explicite les traits distinctifs.

Mais, nous l'avons montré dans la section précédente, cette interprétation par défaut peut être annulée. Que le document ou la section du document concerne une thématique qui se focalise sur une autre entité que celle qui apparaît comme étant plus centrale dans le domaine, et c'est celle-ci qui constitue alors le référent activé préférentiellement. C'est ce qui explique que, dans certains documents du *corpus* professionnel, on ait davantage d'occurrences de *le réseau* pour dénoter *le réseau Erato* que pour dénoter *le réseau routier*. De la même manière, dans certains documents, *le système* renvoie au *système Erato* alors que dans d'autres documents, cette même expression *le système* renvoie au *système Capitoul-2*.

Malgré les évidentes possibilités d'orienter la lecture d'une occurrence particulière, il demeure que, selon la sphère d'activité, un terme réduit recevra préférentiellement une certaine interprétation. La preuve en est que, si l'on inscrit le propos dans une autre activité, la même forme recevra une autre interprétation. Dans des textes de la SEMVAT (responsable au moment de l'étude des transports en commun de l'agglomération toulousaine), on rencontre aussi *le réseau*, utilisé hors situations de reprise anaphorique, sans antécédent textuel. Mais, pour la Semvat, *le réseau* n'est aucunement *le réseau routier* ou *le réseau Erato*, c'est *le réseau de transports en commun*. Rien n'interdit *a priori* que, dans d'autres domaines encore, *le réseau* puisse être, par défaut, le réseau d'eau potable, le réseau électrique, le réseau informatique, etc.

Autre exemple, dans le domaine couvert par le *corpus* sport-loisir, une forme comme *l'activité*, sans plus de précision, renvoie à la notion d'*activité thermique*. Mais dans la sphère de la volcanologie, *l'activité* est volontiers *l'activité volcanique* (merci à A. Josselin-Leray pour cette précision, cf. Josselin-Leray et Roberts 2004, pour une analyse des termes de volcanologie).

Formulation elliptique et interprétation par défaut se joignent pour faire reposer sur le domaine de connaissance une part non négligeable de la signification. Celui-ci promet un référent (ou une série de référents) au rang d'élément cognitivement saillant, quasiment constamment accessible, ou tout au moins facilement récupérable à la moindre évocation de la tête du terme correspondant. Il est donc clair que la formulation elliptique ne peut « fonctionner » convenablement que tant que les textes circulent d'un expert à un autre expert *de la même sphère d'activité*, c'est-à-dire d'un locuteur pourvu de connaissances approfondies, non seulement du domaine, mais de ses habitudes langagières, à un autre locuteur en tous points similaire. Dès que les textes sortent de ce circuit où la connaissance partagée autorise de plus en plus de sous-spécifications et d'implicite, les formes que nous reconnaissons comme termes réduits tendent à s'opacifier.

L'emploi de termes réduits sans antécédent textuel peut aussi être vu comme une marque de la connivence qui existe entre locuteurs d'une même sphère d'activité. Leur expertise, la connaissance des habitudes langagières du domaine, en un mot les connaissances partagées par ces locuteurs, sont autant de facteurs qui favorisent ce type d'emploi de termes réduits. Ces connaissances d'arrière-plan supposées communes sont ainsi décelables « en creux » dans le texte : elles concernent tout ce qui n'est pas explicitement exprimé et qu'il faut cependant mobiliser dans l'élaboration du discours.

4. CONCLUSION

Nous avons dans cet article essayé de montrer que certaines occurrences de formes simples ayant la valeur d'un terme complexe permettent d'apprécier des différences de statuts des notions d'un domaine. En nous basant sur la comparaison quantitative de ces formes employées comme reprise anaphorique ou employées sans antécédent textuel, nous avons mis en évidence que, contrairement à ce à quoi on pourrait s'attendre, les locuteurs d'un domaine privilégient les formes du type *le N* sans antécédent textuel. Nous y voyons l'indice du caractère central de la notion dans le domaine, grâce à quoi une interprétation par défaut de la forme est possible.

Ce type d'analyse éclaire selon nous d'un point de vue différent la recherche terminologique, qui converge avec la proposition de terminologie textuelle de Bourigault et Slodzian (1999). Si l'on peut, dans une certaine mesure, s'affranchir des textes spécialisés pour certaines études (par exemple la morphologie des termes ou les procédés néologiques), l'analyse des cotextes d'emploi des termes sous leurs diverses formes permet d'atteindre un niveau qui n'est plus seulement celui de la notion brute ou de sa dénomination, mais le niveau du discours. Et celui-ci, par les traces d'opérations langagières qu'il recèle, reflète des opérations cognitives, des traitements différenciés de certaines notions. C'est là une voie d'accès nouvelle aux domaines spécialisés, à leurs systèmes linguistiques et notionnels.

BIBLIOGRAPHIE

- Apothéloz (D.), 1995a : « Nominalisations, référents clandestins et anaphores atypiques », dans *Tranel* 23, pp. 143-173.
- Apothéloz (D.), 1995b : *Rôle et fonctionnement de l'anaphore dans la dynamique textuelle*. Droz, Genève.
- Bourigault (D.), Fabre (C.), Frérot (C.), Jacques (M.-P.) et Ozdowska (S.), 2005 : « Syntex, analyseur syntaxique de corpus », dans *TALN'2005*, vol. II, 6-10 juin 2005, Dourdan, pp. 17-20.
- Bourigault (D.) et Slodzian (M.), 1999 : « Pour une terminologie textuelle », dans *Terminologies nouvelles* 19, pp. 29-32.
- Charolles (M.), 2002 : *La Référence et les Expressions référentielles en français*, Ophrys, Gap/Paris.
- Collet (T.), 2000 : *La Réduction des unités terminologiques complexes de type syntagmatique*, Ph. D. thesis, Université de Montréal.
- Corblin (F.), 1987 : *Indéfini, Défini, Démonstratif*, Droz, Genève, Droz.
- Corblin (F.), 1995 : *Les Formes de reprise dans le discours. Anaphores et Chaînes de référence*, Presses universitaires de Rennes, Rennes.
- Cornish (F.), 2001 : « Anaphora, Text, and the Construction of Discourse : A Practical Application », dans Degand (L.), Bestgen (Y.), Spooren (W.) et van Waes (L.), édit., *Multidisciplinary Approaches to Discourse*, Amsterdam & Münster, Stichting Neerlandistiek & Nodus Publikationen, pp. 111-122.
- Cornish (F.), 2003 : « The roles of (written) text and anaphor-type distribution in the construction of discourse », dans *Text*, 23(1), pp. 1-26.
- de Mulder (W.), 1994 : « Déterminants, cohérence et raisonnement par défaut », *Travaux de linguistique*, 29, pp. 93-105.
- Freixa (J.), 2002 : *La variació terminològica. Anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient*, tesi doctoral, Universitat de Barcelona.
- Jacques (M.-P.), 2003a : *Approche en discours de la réduction des termes complexes dans les textes spécialisés*, doctorat nouveau régime, Université Toulouse II-Le Mirail.
- Jacques (M.-P.), 2003b : « Repérage de termes réduits : intérêt et limites de l'analyse distributionnelle », *Cinquièmes rencontres Terminologie et Intelligence Artificielle, TIA-2003*, 31 mars-1^{er} avril, Strasbourg, LIIA-ENSAIS, pp. 135-144.
- Josselin (A.), Roberts (R. P.), 2004 : « La définition des termes dans les dictionnaires de langue générale : analyse de quelques exemples tirés du domaine de la volcanologie à la lumière d'un corpus de vulgarisation », *actes du 72^e congrès de l'ACFAS (colloque OLST)*, 13-14 mai 2004, Montréal, Canada. [à paraître aux Presses de l'Université d'Ottawa].
- Kleiber (G.), 1986 : « Adjectif démonstratif et article défini en anaphore fidèle », dans David (J.) et Kleiber (G.), édit., *Déterminants : syntaxe et sémantique*, Paris, Klincksieck, pp. 169-185.
- Reichler-Béguelin (M.-J.), 1988 : « Anaphore, cataphore et mémoire discursive », *Pratiques*, 57, pp. 15-43.

LA TRADUCTION TECHNIQUE : LE TEXTE SOUS L'EMPIRE DE L'EXTRATEXTUEL

Mathilde Julie Livia Fontanet

département de traduction et de traductologie français
École de traduction et d'interprétation (ETI)
Université de Genève
Genève, Suisse

Résumé : *La présente communication a pour objet de contribuer à éclaircir le statut singulier de la traduction des textes techniques, au sens étroit du terme. Nous nous proposons d'énoncer les particularités de ce type de traduction en y caractérisant l'original, sa fonction, son auteur, le processus de traduction, ainsi que les aspects formels et la portée référentielle du texte traduit, puis nous détaillerons les difficultés propres à cette activité.*

Mot-clés : *traduction technique, ambiguïtés syntaxiques, repérage des termes, découpage sémantique*

1. LE TEXTE TECHNIQUE

1.1. Définition

Le mot « technique » est pris ici au sens étroit, correspondant à la première acception qu'en donne le *Trésor de la langue française*¹ : « qui concerne les applications de la science [et] de la connaissance scientifique ou théorique, dans les réalisations pratiques, les productions industrielles et économiques ». Dans la présente communication, relèvent ainsi des « textes techniques » les manuels de référence, les brochures d'entretien, les modes d'emploi, les spécifications techniques et tout autre document à vocation opérationnelle.

Nous nous appuyerons en outre sur une définition strictement fonctionnaliste du « texte technique » dans la mesure où nous le caractérisons par sa seule mission, sans considérer comme déterminant qu'il soit le plus souvent une manifestation d'un discours spécialisé. Selon nous, le texte technique est purement utilitaire en ce sens qu'il vient répondre au besoin d'informations d'un lecteur désireux de mener à bien une opération pratique (montage d'une machine, application d'une procédure de mise en service, production d'une pièce répondant à certaines exigences, etc.). Ainsi, dans notre perspective, s'il est généralement riche en termes techniques et s'il s'énonce volontiers en

¹ *Trésor de la langue française: dictionnaire de la langue du XIX^e et du XX^e siècle*, (1789-1960), 16 volumes, Paris, CNRS, 1971-1994.

langue spécialisée, les termes et la langue spécialisée n'en sont pas les caractères définitoires.

1.2. La fonction du texte technique

Le texte technique vise à transmettre des données objectives (quantifiées, qualifiées et ne relevant ni de l'opinion, ni du goût) à des lecteurs (ou « utilisateurs ») comptant agir efficacement dans la sphère extralinguistique. À la fois miroir et voie d'accès, il entretient ainsi une relation d'immédiateté avec la réalité, dont il doit être le reflet direct et dans laquelle il doit permettre d'exercer un effet direct. En cela, il se distingue du texte scientifique, qui a d'ordinaire pour vocation d'apporter des connaissances ou de présenter des éléments à l'appui d'une théorie. Si les deux types de discours se doivent d'être précis, concrets, logiques et univoques, le texte scientifique possède une dimension rhétorique et argumentative qui reste étrangère au discours technique.

1.3. Le rapport du texte technique à son destinataire

Le texte technique n'a pas à justifier sa présence, ni à promouvoir sa propre diffusion. Nul besoin pour lui de se frayer une voie vers le public, de chercher à s'imposer parmi les autres discours. Il vient répondre à une demande préexistante, sans avoir ni à la susciter, ni à l'entretenir : le technicien consultant son manuel de procédures n'a pas d'autre option que de s'y référer. Aussi jamais le texte technique ne s'essaie-t-il à la séduction, l'argumentation ou l'émotion. Jamais il ne recourt à l'humour, l'identification ou l'esthétique. Il reste dense de sens, se veut intégralement dénotatif et affranchi de toute portée connotative.

1.4. L'auteur du texte technique

Autre particularité, le texte technique ne se présente pas comme le produit d'un « auteur ». Il semble bien plutôt émaner directement de la réalité technique, avoir été dicté par une forme de logique universelle, sans avoir transité par une quelconque subjectivité. La plupart des autres textes de spécialité laissent entendre une voix, entrevoir un principe humain à leur origine. Le discours scientifique, par exemple, qui procède souvent de la démonstration, suit une ligne de pensée logique et déductive qu'il met en avant pour attester sa crédibilité, et manifeste ainsi en tout temps une intentionnalité. De même, le texte juridique, qui marque son appartenance à une élite par une syntaxe et des tournures archaisantes, permet à son auteur de se positionner socialement. Le texte technique, lui, trouve sa force dans sa neutralité et son objectivité.

1.5. La forme du texte technique

Parce qu'il n'a d'autre vocation que d'informer et que l'information qu'il communique ne saurait prêter à controverse, le texte technique se focalise sur son seul sens, subordonnant tout aspect formel à la seule efficacité du message. Selon Claude Bédard, les textes qui nous intéressent présentent un niveau de langue qu'il qualifie de « niveau usuel », appartiennent au registre neutre (parce qu'ils ne déploient aucun effet stylistique) et marquent le « degré zéro de l'écriture » (Bédard 1986 : 166-168).

Si la forme du texte technique ne cherche pas à colorer le sens, mais uniquement à le mettre à nu, sans jamais le détourner du chemin de la stricte signification, et si elle tend à se faire oublier *per se*, elle n'en reste pas moins le vecteur essentiel du message. Sa qualité se mesure à sa précision, sa clarté et sa concision.

Par voie de nécessité, le texte technique est le plus souvent exprimé dans une « langue spécialisée » (ou « langue de spécialité »), que Lerat définit comme procédant de « l'usage d'une langue naturelle pour rendre compte techniquement de connaissances spécialisées » (Lerat 1995 : 21) et comptant parmi ses expressions caractéristiques les termes, les formules et le « vocabulaire de soutien » (Lerat 1995 : 3). L'usage de la terminologie y est dicté par des contraintes d'univocité (facteur à la fois de clarté et de précision) et de concision.

2. LA TRADUCTION DU TEXTE TECHNIQUE

2.1. La fonction du texte traduit

Le texte traduit a une fonction assimilable à celle de l'original. Ce sont les mêmes informations qu'il vise à transmettre, pour permettre d'exécuter les mêmes gestes et de mener à bien les mêmes opérations. Tout comme l'original, il se destine avant tout à un « utilisateur » et se caractérise par sa nécessaire immédiateté avec la « réalité ».

2.2. Le rapport du texte traduit à l'original

Le texte traduit entretient un rapport tout à fait paradoxal avec le texte original. Comme son centre de gravité se situe en quelque sorte en dehors de la langue, dans la seule « réalité technique », le traducteur peut, si celle-ci l'exige, s'écarter librement du « dire » de l'original, sans même nécessairement chercher à s'appuyer sur le « vouloir dire » de l'auteur : il doit communiquer ce que le texte « devrait dire » pour rester en adéquation avec sa portée extralinguistique.

À l'ordinaire, l'original a une valeur absolue et le texte traduit a une valeur relative. Dans le cadre de la traduction technique, cependant, seul le monde extralinguistique a une valeur absolue, celle de l'original comme de sa traduction restant en tout temps relative. Aussi le principe de la fidélité à l'égard de l'original y est-il atténué : si l'original est mal rédigé ou s'il comporte des erreurs, le traducteur a toute latitude pour intervenir pour réorganiser la forme et corriger le sens. À titre d'exemple, il importe peu que la traduction d'un mode d'emploi emboîte le pas au texte original. S'il s'agit de monter une bibliothèque, l'essentiel est que le lecteur reçoive des informations correctes, claires et précises. En fin de compte, la qualité du travail du traducteur ne se mesurera pas à sa fidélité à l'égard de l'original, mais bien plutôt au temps que mettra l'utilisateur pour exécuter le montage.

C'est précisément la perspective qu'adopte Barbara Folkart, en soulignant que le discours technique est la manifestation la plus absolue de la fonction référentielle de la langue. Selon elle (Folkart 1984 : 229-230), la trajectoire qui s'impose au traducteur technique part de la langue source et passe par le référent extralinguistique pour arriver

à la langue cible, quitte à court-circuiter le texte source pour partir directement du référent : la seule contrainte formelle qu'il doit s'assigner est de maintenir le caractère transparent du texte pour garantir sa relation d'immédiateté avec la réalité qu'il décrit.

2.3. La voix de l'auteur et celle du traducteur

Dans une perspective générale, s'élevant contre la croyance que la « voix de l'auteur » peut être totalement préservée dans une traduction sans que celle du traducteur se laisse entendre, Folkart (1991 : 395-396) avance que « la voix qui parle dans la traduction [...] est une somme de différences plus ou moins minuscules, plus ou moins hétéroclites, une série de dissonances intermittentes et fugaces analogues aux interférences acoustiques, battements produits par deux ondes de fréquence rapprochée ». De même, Theo Hermans (2002) fait valoir que la traduction ne saurait être assimilée à une simple réénonciation de l'original et postule que tout texte traduit, loin d'être une représentation transparente du texte source, est, au même titre que tout autre texte rapporté mais à plus forte raison encore, nécessairement pluriel, décentré, hybride et polyphonique : il est imprégné de la présence discursive du traducteur, qui se manifeste par une voix distincte et une position subjective – que Hermans appelle « la voix différentielle du traducteur » en s'inspirant d'une expression de Folkart. Selon nous, la traduction technique constitue une exception à cet égard. Le total assujettissement du texte à sa valeur référentielle évince tant l'auteur que le traducteur de leur matérialité discursive et fait taire la voix de l'un comme de l'autre.

Hermans estime en outre que la nature polyphonique de la traduction découle aussi du fait que son discours renvoie non seulement au texte source, mais aussi (par soumission ou par transgression) à des normes de traduction, qui viennent « filtrer » les textes que produisent les traducteurs. Nous estimons comme lui qu'une norme prévaut nécessairement dans toute forme de traduction. Dans le cas de la traduction technique, toutefois, celle-ci renvoie à nouveau à la réalité extralinguistique : le mode de dire est subordonné au mode de faire. Ainsi, dans une procédure d'entretien comme dans un mode d'emploi, l'ordre d'énonciation des éléments d'information devra suivre la chronologie des gestes préconisés.

2.4. La forme du texte traduit

La forme du texte traduit, comme celle du texte original, a pour seule fonction de recentrer le texte sur son sens. Aussi doit-elle également se caractériser par sa concision, sa simplicité, sa cohérence et sa clarté.

2.5. Le processus de traduction

Le processus de traduction technique se caractérise par un calibrage répété des incertitudes (pour se protéger des *a priori*, des évidences illusives, et définir le travail nécessaire pour comprendre réellement le texte) et par la nécessité de procéder à des contrô-

les itératifs de cohérence (pour veiller à la cohérence, aussi bien interne qu'externe², de l'interprétation donnée à l'original, puis à la cohérence du texte traduit). Dans la phase de sémiologie, le traducteur doit :

- définir ses incertitudes (car c'est à cette seule condition qu'il pourra les gérer) ;
- déterminer le niveau de compréhension qui lui est nécessaire (il n'a par exemple pas besoin de comprendre le fonctionnement du tube cathodique du seul fait que le mot télévision figure dans un texte) ;
- procéder au repérage des unités sémantiques et terminologiques ;
- entreprendre les recherches terminologiques et documentaires qui s'imposent pour lever les incertitudes ;
- contrôler la cohérence (interne comme externe²) de la compréhension du texte original (contrôle des cohérences interne et externe).

Dans la phase d'onomasiologie, le traducteur doit :

- déterminer le degré de précision que requiert le destinataire ;
- trouver la terminologie adéquate pour restituer (ou du moins transmettre) le sens ;
- opter pour une forme adéquate (univoque, claire et concise) ;
- contrôler le sens (contrôle de la cohérence, tant interne qu'externe² du texte traduit) et la forme (contrôle orthographique, syntaxique et terminologique).

Le traducteur technique cherche d'abord à extraire tous les éléments d'information que recèle l'original. Il sollicite aussi ses connaissances générales et spécialisées, de même que son savoir terminologique. Si ceux-ci sont insuffisants, il entreprend des recherches documentaires et terminologiques. À cet effet, il s'adresse à des spécialistes ou consulte des encyclopédies, des glossaires et d'autres ouvrages de référence. Il peut également s'appuyer sur des *corpus* de textes pertinents. De fait, la proximité entre discours technique et réalité extralinguistique fait que la connaissance de cette dernière peut passer par la consultation des textes fiables qui s'y rapportent. Les *corpus* pertinents, dans la mesure où ils sont assimilables à des manifestations directes de la réalité technique, sont de sûres références pour établir tant le sens que la terminologie. Ils permettent au traducteur de trouver les termes dans leur environnement « naturel » et contribuent donc utilement à la résolution de plusieurs problèmes propres à la traduction technique.

² Par « cohérence interne », nous entendons la compatibilité mutuelle des divers éléments sémantiques du texte. Par « cohérence externe », nous entendons la compatibilité entre chaque élément sémantique du texte et la réalité extralinguistique.

3. LES DIFFICULTÉS INHÉRENTES À LA TRADUCTION TECHNIQUE

3.1. Le repérage des termes

On se méprend souvent sur les difficultés inhérentes à la traduction technique, qu'on tend à ramener à un problème de recherche terminologique, alors que celle-ci peut le plus souvent se mener à bien sans problème. La difficulté première réside en fait souvent dans l'identification des termes, qu'il faut nécessairement repérer avant d'entreprendre de résoudre les problèmes qu'ils entraînent. Or, comme le fait valoir Juan Sager (2000 : 47), le profane peut prendre pour un terme ce qu'un spécialiste considère comme un mot de la langue générale et, inversement, prendre un terme pour un mot ordinaire. Il est parfois difficile de distinguer la langue générale et la langue spécialisée, car celle-ci utilise volontiers des termes spécialisés entremêlés de mots ordinaires. Si les composés savants formés à partir de racines grecques ou latines sont relativement faciles à reconnaître, les termes syntagmatiques, plus fréquents du fait de leur flexibilité formelle et sémantique et de leur productivité (Kocourek 1991 : 151), posent davantage de difficultés. De plus, certains mots du lexique général peuvent prendre une acception particulière en langue spécialisée.

En fait, le plus ardu peut être non pas de trouver la traduction des termes, mais de les définir dans la chaîne syntagmatique et d'établir comment celle-ci s'articule. Prenons l'exemple du titre d'un document du CERN (Organisation européenne pour la recherche nucléaire),

Supply of fine blanked austenitic steel collars for the cold masses of the LHC superconducting dipole magnets.

Pour mener à bien sa tâche, le traducteur doit tout d'abord établir des liens entre les mots, en appairer certains et envisager toutes les possibilités avant d'en écarter le plus possible. De fait, quelles sont les unités à prendre en compte ? S'agit-il de *steel collars* et, dans ce cas, *austenitic* vient-il qualifier *collars* ou *steel*. De même, l'adjectif *fine* s'associe-t-il à *blanked* ou qualifie-t-il directement *collars* ou *steel* ?

3.2. Les différences de plages sémantiques entre les langues

Une autre difficulté réside dans la différence qui peut exister entre le découpage sémantique prévalant dans les deux langues de travail. Par exemple, les trois phrases qui suivent ont été tirées d'un même texte sur la foudre (trouvé sur le site de la BBC³). Or, le terme *lightning*, qui apparaît trois fois dans l'original, ne saurait être traduit par le même mot dans les trois cas :

- 1) *We now know that lightning is an electrical discharge from cumulonimbus clouds.*
- 2) *First, the lightning travels very rapidly towards the Earth, but is quite faint.*

³ Texte trouvé à l'url : <http://www2.thny.bbc.co.uk/weather/features/weatherbasics/lightning.shtml> (consulté le 30 septembre 2005)

- 3) *As lightning is seen and thunder heard, lightning travels at the speed of light, and thunder at the speed of sound.*

Dans la première phrase *lightning* est défini en tant que phénomène, celui d'une décharge électrique provenant d'un cumulonimbus. C'est le phénomène de la foudre.

Dans la deuxième phrase, il est indiqué que le (ou la) *lightning* se déplace très rapidement vers la terre, mais n'est que faiblement lumineux. À l'évidence, le traducteur ne saurait donc opter pour « foudre ». Une recherche documentaire permet d'établir qu'il ne s'agit encore que d'une décharge atmosphérique, autrement appelée « traceur descendant » ou « précurseur » qui est un canal ionisé se développant du nuage vers le sol.

Dans la troisième phrase, enfin, il est fait référence à la manifestation visuelle de la foudre – autrement dit, à l'éclair (qui s'oppose ici au tonnerre, la manifestation auditive).

3.3. L'impossibilité de s'appuyer sur des connaissances préalables

Lors de la lecture d'un texte standard, la préconnaissance du domaine permet au destinataire de procéder à des interprétations (le plus souvent spontanées) qui viennent compenser des ambiguïtés du texte. Un certain flou grammatical n'est ainsi souvent pas même perçu comme tel, car le champ des possibilités est considérablement réduit par l'élimination spontanée de toutes les interprétations en rapport d'incohérence avec la réalité extralinguistique. Inconsciemment, tout lecteur (et tout traducteur) fait intervenir une part importante de déduction et d'inférence pour établir, à partir d'un énoncé quelconque, le sens qu'il doit transmettre.

Ainsi, dans la phrase suivante, la syntaxe anglaise, parce qu'elle autorise à ne pas donner la marque du pluriel au premier élément d'un nom composé, entraîne une incertitude pour quiconque ne connaît pas le contexte.

Magnet training is the critical stage.

Ici, aucune ressource terminologique, ni même documentaire ne pourra permettre de conclure s'il convient de traduire la phrase par « L'accommodation *de l'aimant* est l'étape délicate » ou « L'accommodation *des aimants* est l'étape décisive ».

Un autre problème se pose dans l'expression suivante :

Our klystron cannot be compared with other premium-grade klystrons.

Il est fait référence à « d'autres klystrons ». Or, selon la grammaire anglaise, il peut soit s'agir d'autres klystrons, qui sont *eux aussi* de qualité supérieure, soit d'autres klystrons qui se trouvent par ailleurs être de qualité supérieure. À nouveau, seule une connaissance du contexte sera susceptible de lever l'ambiguïté.

Enfin, outre le problème bien connu des « faux amis », le traducteur doit faire preuve d'une vigilance toute particulière. Pour reprendre un exemple d'une traduction anglais-français, la phrase suivante risque fort d'être mal traduite par tout francophone ne cherchant pas à se représenter la situation :

The purpose of combustion is to rotate the rotor around the chamber.

Sachant que la préposition *around* signifie le plus souvent « autour de », la tentation sera grande de traduire le texte par « La combustion a pour effet de faire tourner le rotor *autour* de la chambre » – une traduction fautive, car le rotor se trouve en l'occurrence à *l'intérieur* de la chambre.

4. CONCLUSION

En raison du statut particulier du texte technique, qui se doit d'être un reflet de la réalité extralinguistique, lors de l'interprétation du texte original, le traducteur ne devra pas s'attacher à restituer « ce qu'a voulu dire » l'auteur, mais « ce qu'il aurait dû » écrire pour que le sens corresponde à cette réalité extralinguistique. Cette polarisation du texte sur « l'extratextuel » a des incidences très sensibles sur les principaux paramètres de la traduction : le filtrage de l'information est ainsi tout autre que dans une situation où il importe de reproduire des effets stylistiques (qui exigent de se concentrer sur le jeu des évocations et sonorités du texte), ou de reproduire un contenu sémantique empreint de subjectivité (car il convient alors de communiquer le message précis que veut transmettre l'auteur).

BIBLIOGRAPHIE

- Bédard (C.), 1986 : *La Traduction technique – principes et pratique*, Montréal, Linguattech.
- Folkart (B.), 1991 : *Le Conflit des énonciations : traduction et discours rapporté*, Québec, Balzac.
- Folkart (B.), 1984 : « A Thing-bound Approach to the Practice and Teaching of Technical Translation », dans *Meta*, n° 29, 3, pp. 229-246.
- Hermans (T.), 2002 : « Paradoxes and apories in translation and translation studies », dans Riccardi (A.), éd., *Translation Studies – Perspectives on an Emerging Discipline*, Cambridge, Cambridge University Press, pp. 10-23.
- Kocoureck (R.), 1991 : *La Langue française de la technique et de la science : vers une linguistique de la langue savante*, 2^e éd. augmentée, refondue et mise à jour, O. Brandstetter, Wiesbaden.
- Lerat (P.), 1995 : *Les langues spécialisées*, Paris, Presses universitaires de France.
- Sager (J.), 2000 : « Pour une approche fonctionnelle de la terminologie », dans Thoiron (Ph.) et Béjoint (H.), dir., 2000 : *Le Sens en terminologie*, Lyon, Presses universitaires de Lyon (Travaux du C.R.T.T), pp. 40-60.

LA TERMINOLOGIE D'ENTREPRISE ET SES CONTEXTES D'USAGE

Franco Bertaccini et Alessandra Matteucci

SSLMIT Forlì, Université de Bologne

47100 Forlì (FC), Italie

alessandra.matteucci@gmail.com, bertacci@sslmit.unibo.it

Résumé : *Ce travail de recherche socioterminologique s'est fait à travers la distribution d'un questionnaire ad hoc sur lequel on a construit des modèles de création terminologique à partir du contexte d'usage considéré comme le pivot autour duquel tournent les objets et leur fonction, les termes et les concepts. L'individuation des termes à utiliser s'est faite donc à travers le filtre du contexte d'usage qui nous a permis de choisir l'équivalent correct dans la ou les langues d'arrivée.*

Mots-clés : *recherche socioterminologique, contextes d'usage, objets, fonction, termes, concepts*

1. INTRODUCTION

L'activité de monitoring d'une réalité linguistique en mutation continue, suivant les nouvelles technologies, peut exister seulement à travers une étude flexible de la terminologie en tant que miroir d'une société en évolution dans laquelle les facteurs extralinguistiques semblent être toujours plus importants : « *some terminological phenomena can be better described by using more flexible and powerful structures of concepts* » (Temmeraman dans Kageura (2002 : 20)).

Ce travail de recherche socioterminologique se fonde donc sur deux points fondamentaux : le fait que la langue n'est pas un système abstrait, et qu'elle se concrétise à partir de son emploi de la part de « communautés sociales » qui peuvent l'adapter à de nouvelles exigences expressives, telles que la création de synonymes et variantes intralinguistiques, ou en adopter les facettes déjà existantes propres à un domaine spécifique.

2. LES USAGERS DE LA TERMINOLOGIE D'ENTREPRISE

Si l'on considère le langage comme la forme par excellence d'expression et d'identification sociale des interlocuteurs, on pourrait bien décrire une multitude de réalités communicatives à partir, par exemple, de l'emploi de la langue naturelle en situation de communication « ordinaire », pour en arriver à la description d'une réalité plus spécifique, comme le cas de la communication à l'intérieur de l'entreprise.

Les réflexions sur la systématisation de la terminologie des composants structurels de la motocyclette Aprilia ont donné lieu à une analyse de la « stratification linguistique en milieu de travail », tout en considérant le « réseau social constitué par les individus

ou de petits groupes appartenant à une même catégorie » (Berruto 1974 : 97). Dans notre cas, les groupes classés et étudiés sont les suivants :

- ingénieurs et techniciens (experts) ;
- techniciens, responsables de la production et de la manutention (semi-experts) ;
- administratifs, responsables des ventes et ouvriers (non experts).

Souvent, chacun des groupes mentionnés utilise une terminologie différente sur la base des différentes situations de communication, sans jamais exclure le choix d'une terminologie appartenant à la variété géolectale (dialecte vénète), générant ainsi des cas de *code mixing* et de *code switching*.

Les réponses de nos informateurs nous ont permis en outre de mettre en évidence le fait qu'aux plus bas niveaux d'instruction, correspond souvent une grande présence d'objets et, par conséquent, une plus grande présence de variation. C'est pourquoi l'individuation de la terminologie du domaine doit être étudiée surtout d'un point de vue pratique, qui doit tenir en considération les différentes classes d'utilisateurs qui s'approchent de la terminologie d'une manière unique, surtout à partir des contextes situationnels, du niveau de scolarisation et des contextes d'usage.

2.1. La variation terminologique et ses usagers

Le manque de normes spécifiques pour la formation de la terminologie de la majorité des composants structurels de la motocyclette, a donné lieu à la prolifération incontrôlée d'un jargon d'entreprise, autrement appelé « variation occupationnelle », qui ne se nourrit pas seulement d'éléments linguistiques, mais surtout de facteurs extralinguistiques, voire technologiques. La systématisation de ce genre de terminologie devrait adopter une solution intermédiaire pour la gestion des connaissances occupationnelles basée sur une plus stricte collaboration entre les interlocuteurs-employeurs de plus bas niveaux et les experts de plus haut niveau tels que les ingénieurs, pour la construction d'une *shared knowledge* du métier puisque « *terminologies are deliberate creations, [...] and the unique position of terminology is characterised by the combination of two contradictory factors, i.e. the quest for systematicity and flexibility* » (Kageura 2000 : 19).

Se limiter à la systématisation des termes de la langue standard correspondrait à une fausse représentation d'une réalité terminologique qui se fonde sur d'anciennes racines linguistiques, même si limitées dans l'espace et dans la géographie, mais aussi sur la créativité des locuteurs et leur méthode d'apprentissage du métier. Au niveau intralinguistique, l'approche variationniste prend en considération non seulement la terminologie standardisée, mais elle se focalise surtout sur ce qu'on a déjà appelé « jargon d'entreprise » ou « variation occupationnelle », une langue de spécialité qui se forme en premier lieu sur les « mots » de locuteurs qui peuvent connaître très bien leur métier, mais qui quelquefois manquent d'instruments linguistiques pour la création correcte d'une terminologie professionnelle.

L'oralité, qui se distingue en tant que trait primaire du plus haut niveau de variété, c'est-à-dire le dialecte, devient source inépuisable de vieux et nouveaux termes qui vont compléter l'éventail des usagers : des plus experts aux néophytes comme le montre la représentation graphique suivante.

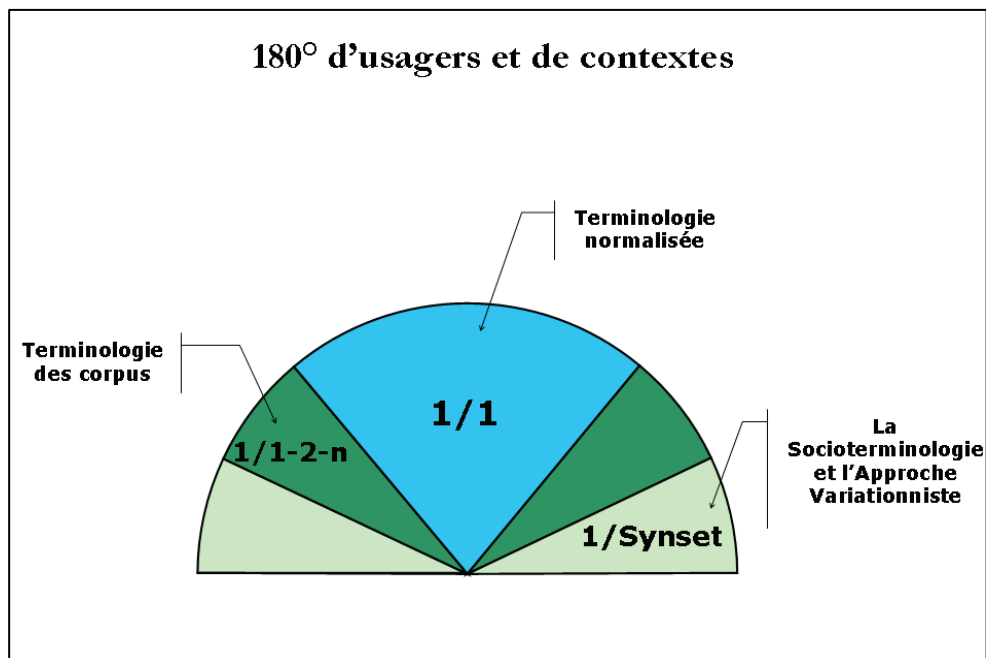


Diagramme 1 – Éventail des usagers de la terminologie

2.1.1. La terminologie normalisée

La terminologie normalisée adopte deux principes fondamentaux pour la systématisation des domaines :

- au niveau intralinguistique, à une notion correspond une dénomination (principe 1 : 1) ;
- au niveau interlinguistique, à une dénomination correspond un équivalent (principe 1 : 1).

Tout en considérant la spécificité de notre recherche, il n'est pas difficile de comprendre que la terminologie normalisée ne réussit pas à décrire de manière ponctuelle la réalité de Aprilia, même si on y fait recours là où elle peut être appliquée de manière immédiate. C'est le cas, par exemple, des composants techniques (rouleaux, visserie, etc.) pour lesquels il existe une terminologie ISO réellement exploitée.

2.1.2. La terminologie des corpora

Le processus de création des *corpora* comporte des choix autoritaires et souvent très subjectifs de la part du terminologue, qui doit se baser sur les fréquences d'usage des termes et, généralement, sur la contribution d'un expert. C'est à partir des données collectées et de leur analyse, que l'on cherche à établir des hiérarchies entre les susdits termes étiquetés comme « termes vedette » et les séries de synonymes et de variantes individuées à l'intérieur des *corpora*.

Ce qui arrive souvent c'est qu'au niveau pratique, les usagers des banques de données organisées sous ce principe ont eux-mêmes du mal à reconnaître les termes vedettes, puisque chaque firme tend à adopter une terminologie qui est pensée pour décrire un produit spécifique. Par exemple, le terme français « support de roue » en italien a comme équivalents les termes *portaruota* (Aprilia), *pedino forcella* (Ducati), mais aussi *gambaletto* ou *portastelo* (autres firmes).

2.1.3. La socioterminologie

De l'analyse des données des *corpora* on a donc développé l'idée de « set de synonymes » (*Synset*) pour laquelle il n'existe pas un véritable terme vedette, puisque la systématisation hiérarchique des termes dépend des applications concrètes de la langue, ce qui permet aux usagers d'entrer dans une base de données ou dans une situation de communication d'une façon pratique, naturelle et immédiate.

Pour la terminologie variationniste, au même concept peuvent donc correspondre plusieurs termes qui forment un véritable *set* de synonymes. De cette manière, on peut arriver à la construction d'une terminologie des besoins concrets pour des objectifs professionnels précis, structurée de manière empirique pour obtenir une communication parfaite à l'intérieur de l'entreprise.

3. La terminologie et les contextes d'usage au sein de l'entreprise

Sur la base des réflexions mûries pendant la période de collaboration avec Aprilia S.p.A, et suite à la réalisation d'un glossaire technique des composants structurels du modèle RSV 1000R 2004, on a pu individuer un « macrocontexte d'usage » de la langue, le siège de Noale (Padoue), qui se compose de plusieurs « microcontextes d'usage » de la langue correspondant aux secteurs dont se compose la susdite firme. Ces espaces, qu'on pourrait appeler aussi « environnements occupationnels » se caractérisent par des dynamiques de communication et par une approche à la terminologie spécifique très différentes, en relation aux compétences individuelles des acteurs sociaux compétents qui constituent le « réseau sociooccupationnel » chez Aprilia.

L'étude des dynamiques communicatives s'est faite à travers la distribution d'un questionnaire de « sociolinguistique en milieu de travail » qui nous a permis de construire des modèles de création terminologique tout en considérant le contexte d'usage comme le pivot central autour duquel tournent les objets et leur fonction, les termes et les concepts. Aux informateurs appartenant aux différents secteurs de la firme, on a en effet demandé de donner les équivalents en anglais et en dialecte vénète, code très exploité chez Aprilia, d'une liste de termes en italien dépourvus de contexte.

Veneto	Italiano	Francese	Inglese
	AMMORTIZZATORE POSTERIORE		
	BOCCHETTONE		
	BUSSOLA CARENA		
	CODONE		
	DADO		
	FODERO GUARNIZIONE		

**Tableau 5 – Extrait du questionnaire de « sociolinguistique occupationnelle » :
la grille des équivalents**

Ce « vide contextuel » s’est révélé l’obstacle principal pour le choix de la terminologie la plus appropriée. Cela met en relief l’importance des contextes d’usage en tant qu’éléments discriminants de la communication écrite ou orale.

3.1. Les microcontextes d’usage

Dans les paragraphes précédents, on a déjà fait référence aux différentes situations communicatives qui peuvent se créer à l’intérieur des espaces de travail. Ces mêmes situations nous ont permis d’individuer trois principales typologies de contextes :

- contexte technique ;
- contexte « sociolinguistique » ;
- contexte graphique.

3.1.1. Contexte technique

Le choix terminologique, au sein d’un contexte technique, est lié principalement à ceux que nous avons appelés « les besoins du métier ». Sur la base des données analysées et tout en excluant la représentation « en triangle » proposée par Wüster, qui serait trop rigide pour décrire les relations réelles qui existent entre objets et leur fonction, les contextes d’usage, les concepts et les termes, il serait intéressant de donner un exemple du processus de désambiguïsation à partir du mot « levier » qui fait partie, sous la forme de ces acceptions spécifiques, de la base de données du glossaire Aprilia.

Le mot « levier » devient terme spécifique à partir d'un ou plusieurs contextes d'usage bien établis qui déterminent son rapport avec les autres objets et notamment avec les autres pièces détachées de la motocyclette Aprilia.

À partir des différents contextes d'usage, on peut individuer des concepts différents qui peuvent décrire des fonctions spécifiques d'objets divers (*i.e.* : levier de frein avant, levier d'embrayage), mais qui peuvent répondre à un seul terme (levier) tacitement désambiguïsé par les interlocuteurs d'un même espace. L'individuation du terme à utiliser selon les exigences de l'interlocuteur se fait à travers le filtre du contexte d'usage qui nous permet même de choisir l'équivalent correct dans la ou les langues d'arrivée.

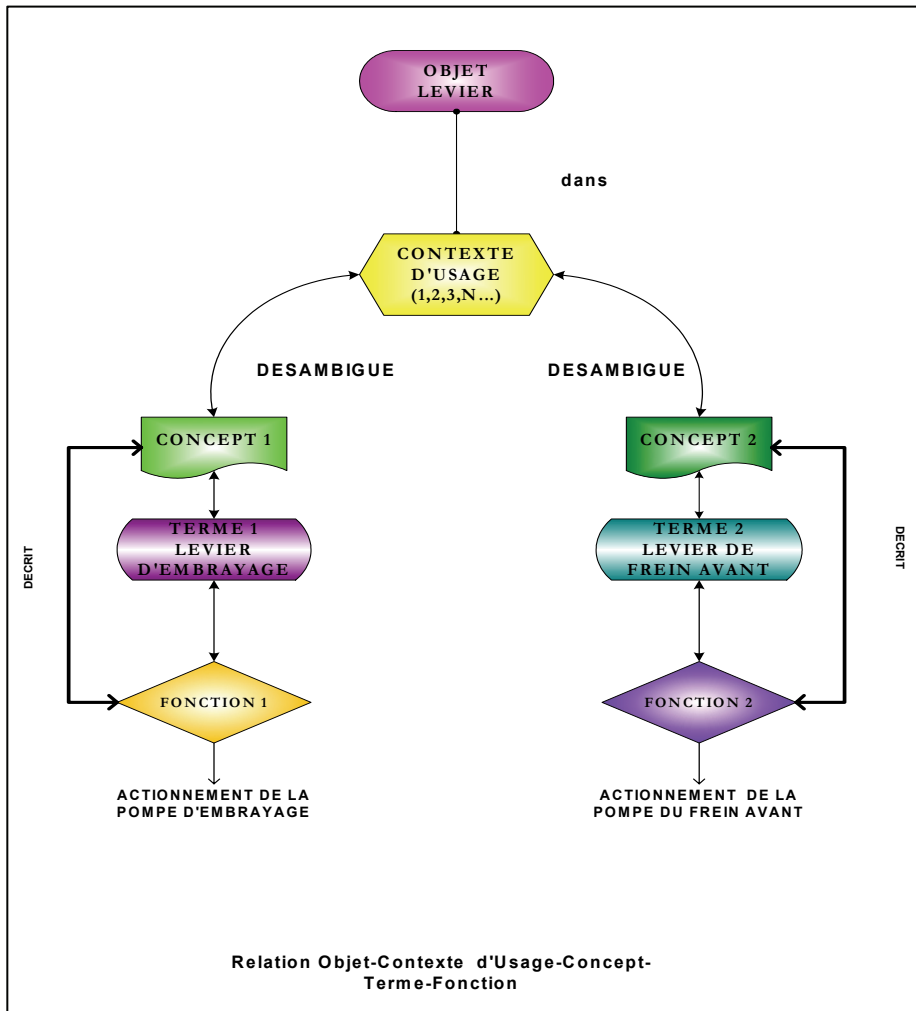


Diagramme 2 – Relations intralinguistiques du mot « levier »

Un autre cas très représentatif du contexte technique : le mot *valvola* en italien. Quel terme devrait-on choisir en français ? Valve, soupape ou papillon ? Le choix dépend du mécanisme interne de cet élément et il nous sert en tant que discriminant pour une bonne traduction terminologique et, pourquoi pas, le choix alternatif mais correct d'une variante ou d'un synonyme.

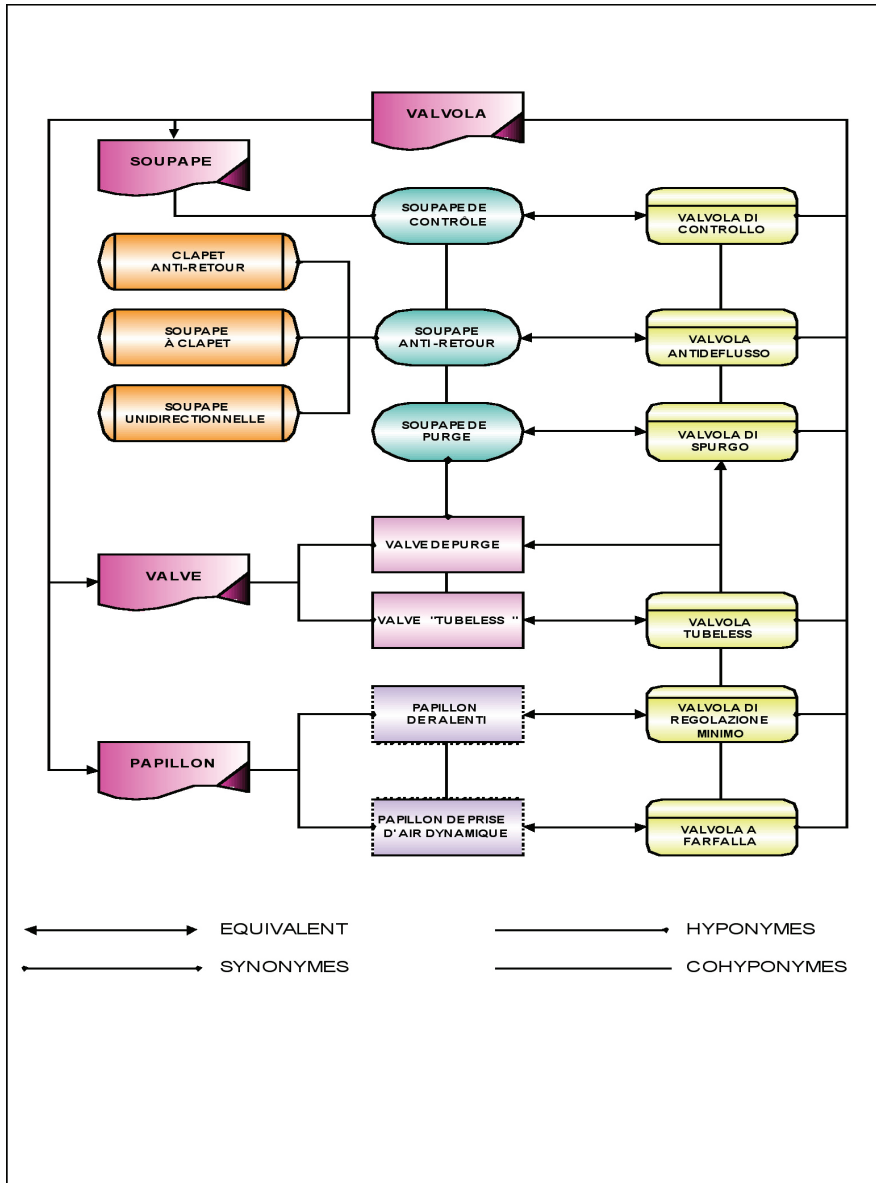


Diagramme 3 – Relations intralinguistiques et extralinguistiques du mot « valvola »

3.1.2. Contexte sociolinguistique

Sur le plan sociolinguistique, le choix terminologique est lié au « statut socioculturel » des locuteurs. Ils utilisent la langue de manière différente sur la base de leurs connaissances personnelles et professionnelles. Tous les niveaux de la langue peuvent ainsi établir un contact et déterminer des choix terminologiques intéressants. Par exemple, le locuteur moins cultivé adoptera la variante dialectale plutôt que le terme italien standardisé.

Terme	ammortizzatore posteriore (amortisseur arrière)	Terme technique
Synonyme	monoammortizzatore	Terme technique Spécifique-standardisé
Variante diastratique	mono (terme abrégé)	Terme technique Spécifique Non standardisé
Variante géolectale	amortisator da drio, amor- ticator par de drio	Terme dialectal

Tableau 6 – Un exemple de variation terminologique

3.1.3. Contexte graphique

Pour ce qui est des éléments qui couvrent la fonction de contexte, il faut considérer aussi les images. La pratique générale de la terminologie prévoit la construction d'arbres conceptuels représentatifs des relations logiques et ontologiques qui lient les termes entre eux, mais pour ce qui est de la systématisation du domaine de la mécanique, ce sont les éclatés qui jouent le rôle des arbres conceptuels.

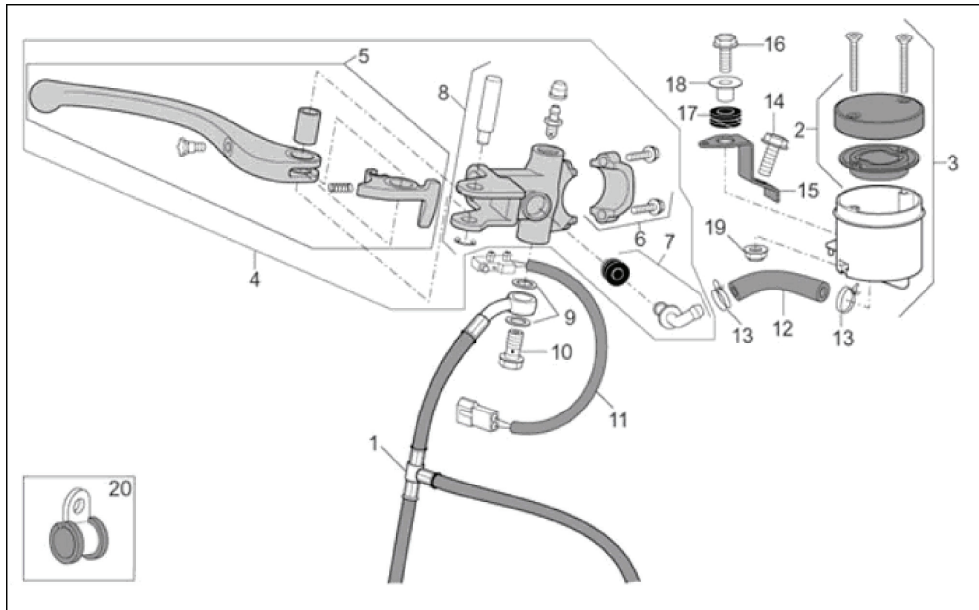


Image 1 – Éclaté du maître-cylindre de frein

Les images des pièces détachées peuvent en effet nous aider dans le processus de désambiguïsation :

- On peut localiser les objets de manière à pouvoir leur attribuer une fonction ;
- Sur la base des images, on peut penser directement au terme correspondant dans la ou les langues connues ;
- On peut aussi composer des définitions de type intensif à partir de la description de ses caractéristiques physiques et de la position par rapport aux autres pièces.

4. CONCLUSIONS

Les contextes d'usage témoignent des dynamiques réelles de la communication et, par conséquent de l'utilisation de la terminologie de notre domaine d'étude, aussi bien à l'écrit qu'à l'oral. Les locuteurs déterminent les contextes et ils attribuent des valeurs et des signifiés différents aux termes dans leur passage de la langue naturelle au « langage occupationnel » ou « jargon d'entreprise ».

Compte tenu du « contexte d'usage » et du « contexte situationnel » on peut affirmer que la terminologie devrait être plus ductile et malléable pour mieux s'adapter à la situation de communication considérée. La communication et l'échange d'expériences professionnelles au moyen de la langue font partie d'un processus sociologique qui ne peut pas et ne doit pas être limité par des procédures de classification trop rigides.

Les variétés, elles aussi, peuvent acquérir, sur la base du contexte d'usage, le statut de langue, puisque l'on peut se trouver dans une situation de communication mixte dans laquelle les usagers peuvent utiliser en même temps tous les outils linguistiques dont ils disposent.

Compte tenu que « *flexible changeability is itself an essential aspect of the structure of a category* » (Geeraerst dans Temmermann 2000 : 129), les termes et leurs possibles variantes contribuent à la description et à la représentation des richesses d'une langue et c'est pourquoi qu'on ne peut pas les négliger.

BIBLIOGRAPHIE :

- Berruto (G.), 1999 : *Sociolinguistica dell'italiano contemporaneo*, La Nuova Italia Scientifica (NIS).
- Berruto (G.), 1994 : *La Sociolinguistica*, Bologna, Zanichelli.
- Kageura (K.), 2002 : *The Dynamics of Terminology. A Descriptive Theory of Term Formation and Terminological Growth*, Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Roy (J.), 2004 : « Approche structurationniste du partage des connaissances terminologiques en milieu industriel et implantation informatique », Université Lyon III, ERSICOM, pp. 1-9.
- Temmermann (R.), 2000 : *Towards New Ways of Terminology Description, The Sociocognitive approach*, Amsterdam, John Benjamins.

CONSTRUCTION COLLABORATIVE D'UN LEXIQUE FRANÇAIS-ANGLAIS TECHNIQUE DANS *ITOLDU* : CONTRIBUER POUR APPRENDRE

Valérie Bellynck^{1et2}, *Christian Boitet*², *John Kenwright*¹

¹ EFPG et cellule TICE, Institut national polytechnique de Grenoble

² GETA-CLIPS-IMAG, Université Joseph Fourier

B. P. 53, 385, rue de la Bibliothèque, 38041 Grenoble CEDEX 9

Grenoble, France

john.kenwright@efpg.inpg.fr ;

valerie.bellynck@imag.fr ;

christian.boitet@imag.fr

Résumé : *La construction collaborative de ressources lexicales libres de droit se heurte à l'impossibilité d'obtenir beaucoup de petites contributions gratuites et volontaires. IToldU est un service web léger qui, lors de sa première année d'utilisation pour l'enseignement de l'anglais technique dans des écoles d'ingénieurs, a permis d'obtenir un peu plus de 17 000 termes anglais dans une vingtaine de domaines, accompagnés de leur traduction en français, correcte à 95 %, et d'exemples d'usage en français, corrects à 85 % environ. IToldU est facile à étendre à d'autres couples de langues. Il reste à trouver comment synchroniser IToldU avec Papillon, une base lexicale multilingue beaucoup plus ambitieuse (<http://www.papillon-dictionary.org>), et à quels autres contextes on pourrait l'étendre pour obtenir de la même façon d'autres types d'informations, comme les fonctions lexicosémantiques.*

Mots-clés : *ressources lexicales, ressources contextuelles, exemples d'usage, construction collaborative, EAO de l'anglais technique*

1. INTRODUCTION

La construction collaborative de ressources lexicales libres de droit se heurte à l'impossibilité d'obtenir beaucoup de petites contributions gratuites et volontaires. *IToldU* (*Interactive technical on-line dictionary for universities*) est un site web pour la construction collaborative d'un lexique bilingue par une petite communauté (par exemple, une promotion d'étudiants) dans un contexte d'apprentissage d'une langue spécialisée sur des domaines techniques. Les contributions sont alors gratuites mais contraintes, car une partie de la note d'anglais est calculée par le site lui-même.

Pour les deux premiers auteurs, l'objectif initial de la réalisation de ce site était de récupérer les lexiques produits, de façon à « abonder » la base de données lexicales multilingue (BDLM) multiusage *Papillon*, en cours de construction. Pour le troisième auteur, enseignant d'anglais et fortement impliqué dans les TICE (Technologies de l'information et de la communication pour l'éducation) à l'INPG (Institut Polytechni-

que de Grenoble), l'objectif était d'améliorer l'enseignement du vocabulaire anglais technique à des élèves-ingénieurs français.

Dans son état actuel, *IToldU* répond bien au second objectif, et pas encore au premier. Mais son usage a fait apparaître un troisième axe intéressant, celui de l'enseignement de la structure des phrases simples de l'anglais et de ce qu'est un « exemple d'usage », car il se trouve que les étudiants ne se contentent pas de copier et coller des phrases contenant les termes qu'ils traduisent, ils préfèrent « forger » des exemples.

Dans ce bref article, nous ne parlerons pas plus du projet *Papillon* et de la BDLM associée, et renvoyons à <http://www.papillon-dictionary.org/> et aux documents disponibles sur ce site pour plus de détails. Dans les trois sections qui suivent, nous présenterons le système *IToldU*, dirons comment il a été utilisé et ressenti par les étudiants et enseignants, et évaluerons les résultats dictionnaires quantitatifs et qualitatifs obtenus au terme de la première année complète d'utilisation.

2. LE SERVICE WEB *ITOLDU*

Les enseignants et les étudiants ne voient dans le site *IToldU* que les fonctionnalités qui leur sont utiles : les étudiants et visiteurs peuvent rechercher, proposer et mémoriser des traductions d'expressions issues de l'anglais technique (ou thématique), les enseignants peuvent obtenir des statistiques, contrôler les productions des étudiants et animer le site par des chasses aux mots. Seul l'enseignant coordinateur peut gérer le site (liste des enseignants, des étudiants, des groupes d'étudiants...).

L'objectif de collecte de données lexicales n'est pas présenté aux étudiants et aux enseignants, qui ne sont conscients que des objectifs affichés par l'enseignant coordinateur :

- Motiver les étudiants à faire du travail « lexical » en dehors des cours ;
- Minimiser la charge de travail supplémentaire des enseignants.

Le résultat est un environnement paramétrable par l'enseignant coordinateur, utilisable par les étudiants pour rentrer les résultats de leur travail lexical et par les enseignants pour diriger puis suivre ce travail.

2.1. Utilisation par les étudiants

Les étudiants doivent rechercher des expressions techniques en anglais, et en proposer des traductions en français. De plus, pour chaque terme, ils doivent proposer (par citation ou forgeage) un exemple d'usage, avec le cas échéant la source (web ou bibliographique) d'où est issu cet exemple. Chaque étudiant a son propre dictionnaire et peut voir tout le dictionnaire de sa promotion.

Le principe utilisé pour motiver les étudiants et autoréguler leurs contributions est simple : l'étudiant commence par rechercher si le terme qu'il veut traiter a déjà été traité par quelqu'un de sa promotion avant de l'introduire. Si oui, et si la traduction et l'exemple lui conviennent, il peut (mais ne doit pas) se l'approprier en l'incorporant dans son dictionnaire personnel. Il peut aussi créer une nouvelle proposition.

Dans les deux cas, *IToldU* lui donne des points. De plus, chaque « appropriation » est considérée comme un « vote » pour l'étudiant dont la proposition a été adoptée, qui gagne des points. Enfin, la publication des meilleurs scores motive les étudiants à participer.

<i>Chercher une traduction</i>	
Mot	moonlighting
Traduction	le travail au noir
Contexte	There is a widespread problem of moonlighting in immigrant populations in Italy which has side-effects on the economy.
Source	invented
Catégorie	business english
Vote	75 % (3/4) <input type="button" value="Charger le mot"/>

Figure 18 : formulaire d'entrée d'un terme dans *ItoldU*

2.2. Utilisation par les enseignants

L'enseignant responsable d'un groupe peut évaluer la participation de ses étudiants à travers des statistiques : *a priori*, plus un étudiant a obtenu de votes, meilleure est sa contribution. Une surveillance minimale sur la qualité des contributions lui permet de détecter toute dérive d'usage du site et donc d'y remédier rapidement.

De plus, tout enseignant peut inciter les étudiants de ses groupes à contribuer un peu plus, de façon un peu ludique, en leur proposant une « chasse aux mots » (*word hunt*).

Un élément important est la prise en compte du fait que les enseignants ont très peu de temps pour suivre les étudiants en dehors des cours, et que l'usage de *IToldU* ne doit pas augmenter leur temps de travail, mais si possible le diminuer. Concrètement, un enseignant a seulement 1 à 2 minutes par cours et par étudiant.

C'est pourquoi, dans la version actuelle, l'enseignant peut marquer une entrée comme fautive, mais ne peut même pas la corriger. C'est à l'étudiant de le faire, et les points afférents lui sont retirés par le système s'il n'a pas corrigé avant un certain délai. On incite bien sûr les étudiants à revoir leur dictionnaire complet, même si, et ils le savent, leurs enseignants n'ont pas le temps de regarder toutes les entrées, et procèdent par échantillonnage.

<input type="checkbox"/> <p>(un)likely => (peu)probable</p> <ul style="list-style-type: none"> • <u>contexte</u> : " For example, if you lost something (jewel) during swimming in the sea, you will be unlikely you find it again. " • <u>source</u> : English lesson • <u>auteur</u> : thierry.finet • <u>vote</u> : 100% (1/1) 	<div data-bbox="864 293 1182 387" style="border: 1px solid black; padding: 5px;"> Erreur : l'enseignant va barrer cet exemple </div>
<input type="checkbox"/> <p>A going away gift => Cadeau de départ</p> <ul style="list-style-type: none"> • <u>contexte</u> : " When someone leave a company " • <u>source</u> : English lesson • <u>auteur</u> : thierry.finet • <u>vote</u> : 100% (1/1) 	
<input type="checkbox"/> <p>Advanced technician => Technicien supérieur</p> <ul style="list-style-type: none"> • <u>contexte</u> : " Professional level " • <u>source</u> : Homework • <u>auteur</u> : thierry.finet • <u>vote</u> : 100% (1/1) 	<div data-bbox="864 524 1150 600" style="border: 1px solid black; padding: 5px;"> Exemple inventé (forgé) </div>
<input type="checkbox"/> <p>assessment => évaluation</p> <ul style="list-style-type: none"> • <u>contexte</u> : " At the end of the year, we will have an assessment of our english level. " • <u>source</u> : Class. • <u>auteur</u> : gaelle.dupuis • <u>vote</u> : 100% (2/2) 	
<input type="checkbox"/> <p>avalanche probe => sonde</p> <ul style="list-style-type: none"> • <u>contexte</u> : " Use avalanche probe to find people under snow " • <u>source</u> : Avalanche safety • <u>auteur</u> : sylvain.bouquet • <u>vote</u> : 100% (1/1) 	
<input type="checkbox"/> <p>Avalanche transceiver => (ARVA) Appareil de Recherche de Victime en Avalanc</p> <ul style="list-style-type: none"> • <u>contexte</u> : " find people under avalanche with avalanche transceiver " • <u>source</u> : Avalanche safety • <u>auteur</u> : sylvain.bouquet • <u>vote</u> : 100% (1/1) 	
<input type="checkbox"/> <p>bank => berge</p>	

Figure 19 : fragment du dictionnaire d'un groupe d'étudiants

2.3. Implémentation

Du point de vue technique, *IToldU* repose sur une base de données *MySQL* (une par promotion), contenant le dictionnaire courant, ainsi que les utilisateurs avec leurs divers droits d'accès. C'est un service web développé en HTML/SQL/PHP, et installé chez un fournisseur gratuit (laposte.net, puis grenet.fr). Il n'y a donc pas de problèmes pour y accéder depuis chez soi, 24 heures sur 24.

Ce système est facile à cloner, à installer sur d'autres sites, et à adapter à d'autres langues, car tous les messages et items de menus sont contenus dans des ressources textuelles éditables sans connaissance informatique particulière.

3. ÉVALUATION DE LA PREMIÈRE ANNÉE D'UTILISATION

3.1. Aspect pédagogique

La version courante du site *IToldU* (<http://opus.grenet.fr/itoldu/ITOLDU>) a été utilisée en grandeur réelle pour la première fois en 2004-2005 par tous les élèves de l'EFPG, une école d'ingénieurs de l'INPG, avec un impact pédagogique positif sensible. Il y avait au total 250 étudiants, répartis sur trois années d'école d'ingénieurs et une de licence professionnelle. Pour l'enseignement de l'anglais, il y avait 17 groupes, 6 enseignants, et 1 enseignant-coordonateur (le 3^e auteur).

IToldU répond déjà assez bien au besoin ressenti par l'enseignant coordinateur de disposer d'un outil informatique améliorant tant l'organisation de l'enseignement que le travail des enseignants et l'apprentissage par les étudiants du vocabulaire technique anglais de spécialité.

Voici par exemple un écran de « chasse aux mots », fonction bien appréciée par les enseignants et par les étudiants. Le premier à trouver un équivalent gagne un point !

perks	Avantages	gaene.dupuis	DUT promo 13
jobless	Au chômage	gaelle.dupuis	DUT promo 13
Employment agency	Agence de placements	thierry.finet	DUT promo 13
nine-to-five job			Ajouter
Hire and fire			Ajouter
Corporate culture			Ajouter
Long-hours culture			Ajouter
Casual Friday			Ajouter
going rate			Ajouter
cash in hand			Ajouter
job with scope			Ajouter

Figure 20 : chasse aux mots (word hunt) préparée par un enseignant

L'utilisation d'*IToldU* a modifié dans le bon sens le comportement des étudiants pendant les cours : ils s'intéressent plus à la prise de notes, et utilisent *IToldU* en dehors des cours comme un support pour leur apprentissage du vocabulaire anglais.

Leur point de vue est maintenant à plus long terme qu'avant, car ils savent qu'ils pourront emporter dans leur vie professionnelle une copie active (installable sur un site web et modifiable) de leur dictionnaire personnel et, s'ils le souhaitent, de tout le dictionnaire constitué par leur promotion.

Il faut cependant noter que les enseignants n'ont pas tous été également convaincus de l'utilité de cette forme de travail, ce qui se reflète dans l'inégalité des contributions des groupes d'étudiants (voir Figure 22 ci-dessous).

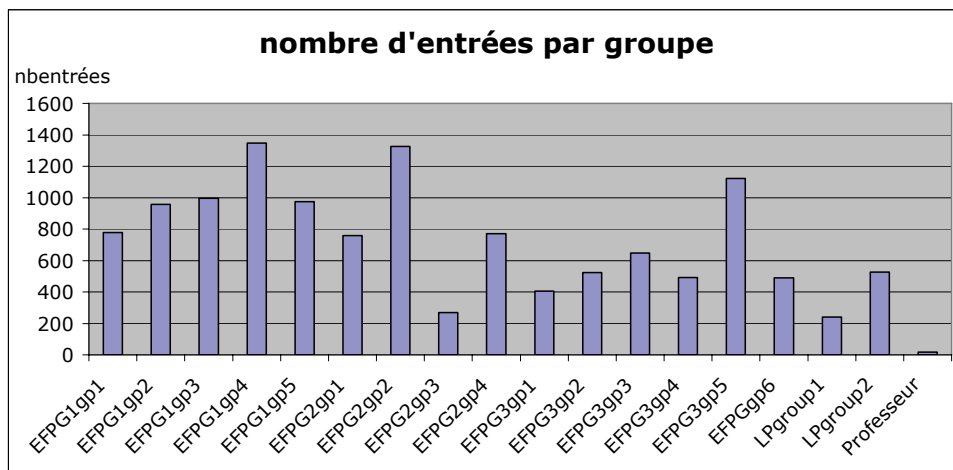


Figure 21 : nombre de contributions par groupes au premier semestre

3.2. Aspect contributif

Le problème de trouver comment motiver les étudiants à contribuer et comment auto-réguler le processus global de contribution est un cas particulier d'un problème général reconnu comme très difficile, celui de susciter des contributions volontaires et gratuites pour « abonder » des bases de connaissances.

Ce problème est difficile, car il n'y a que très peu de spécialistes d'un certain type de connaissances qui acceptent de donner sans contrepartie les connaissances qu'ils ont collectées. Au-delà de quelques contributions de cette nature, qui, même si elles sont individuellement importantes, ne représentent au total qu'une faible fraction des connaissances désirées, il faut faire appel à des non-spécialistes, en grand nombre, contribuant chacun un peu, voire de façon fragmentaire.

Mais, dans les faits, on n'arrive jamais à obtenir des membres d'une communauté d'intérêt des contributions volontaires et gratuites.

Si elles sont volontaires, c'est que les contributeurs « gagnent » quelque chose à le faire, et donc elles ne sont pas gratuites *stricto sensu*. Par exemple, les traducteurs utilisant le site <http://www.yakushite.net/> de Oki Electric mettent des mots dans les dictionnaires car ils utilisent gratuitement des outils d'aide à la traduction dans lesquels les mots ajoutés deviennent quasi instantanément actifs.

Si par contre les contributions sont vraiment gratuites, c'est qu'elles sont « contraintes » d'une certaine manière, la plus discrète et conviviale possible bien sûr. C'est le

cas de *IToldU*, que presque tous les utilisateurs sont « invités fermement » à utiliser, qu'il s'agisse des enseignants ou des étudiants (qui gagnent des points en contribuant).

La méthode retenue dans *IToldU* ne joue pas seulement sur la contrainte, par l'intermédiaire du calcul d'une partie de la note en fonction de l'utilisation (correcte) du site. Elle permet aussi d'établir un esprit de coopération et d'émulation entre les étudiants. D'une part, comme nous l'avons vu, les étudiants coopèrent en « votant » pour ceux dont ils adoptent des entrées. D'autre part, le système présente un « tableau d'honneur » des étudiants qui ont le plus contribué. Enfin, les « chasses aux mots » donnent lieu à une saine et ludique émulation.

4. ÉVALUATION DICTIONNAIRIQUE DE LA 1^{RE} ANNÉE DE CONTRIBUTION

4.1. Aspect quantitatif

En un semestre, environ 12 000 entrées anglais-français ont été mises par les étudiants dans *IToldU*, avec environ 8 000 « contextes d'usage ».

À la fin de l'année universitaire, *IToldU* contenait 17 062 entrées anglais-français, et presque autant de contextes d'usage (16 905, seuls 157 articles n'en avaient pas).

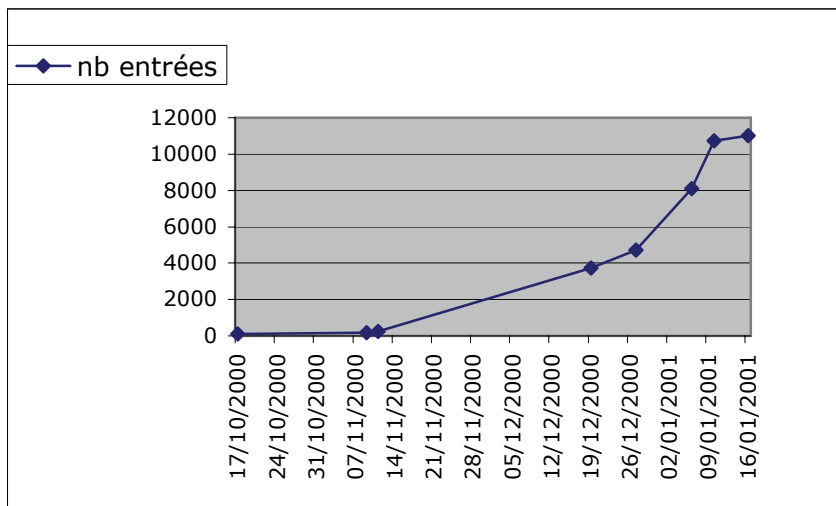


Figure 22 : évolution du nombre d'entrées au premier semestre

4.2. Aspect qualitatif

Le 2^e auteur a revu rapidement l'ensemble des contributions, et 10 % environ en détail, en les corrigeant. Si on met à part les erreurs venant de problèmes de saisie des diacritiques sur le web, on peut dire que les traductions françaises des termes anglais sont presque toutes correctes. Par contre, 15 à 20 % des contextes d'usage ne sont pas des exemples d'usage. Voici quelques détails sur ces deux types de contribution :

4.2.1. Traductions

95 % des traductions nous semblent correctes. Un point intéressant est qu'environ 30 % seulement des termes anglais choisis par les étudiants concernent un champ lexical technique « pur », c'est-à-dire (ici) un des champs liés à leurs études (fabrication de pâte à papier, de papier, de carton, traitement des couleurs, encres, rhéologie...), alors que 70 % concernent des champs « paratechniques » (pour eux), tels que les affaires, la recherche d'emploi... et aussi l'anglais général.

4.2.2. Des « contextes d'usage » vers les « exemples d'usage »

Ce qui concerne les « contextes » mérite quelques commentaires. Dans l'esprit des enseignants, il devait s'agir de citations de phrases dans lesquelles les termes anglais avaient été rencontrés. Mais plusieurs choses inattendues se sont produites.

Certains étudiants ont compris qu'il s'agissait d'un raffinement du « domaine » choisi dans la liste fermée proposée par *IToldU*. On trouve ainsi :

5024	opportunity	possibilité, débouché	society
5025	to put up	ériger, construire	society
5026	to fulfill	accomplir, réaliser	society
5027	fulfilling	profondément, satisfaisant	society
15009	gas-fired	chauffé au gaz	used in paper mill

D'autres ont compris qu'on leur demandait des définitions. On trouve ainsi :

15049	a wind mill	une éolienne	an energy-producing facility
15065	a light bulb	une ampoule électrique	energy-related equipment
4632	TCF (totally chlorine free)	sans chlore	stade de blanchiment

Le coordinateur a alors dit qu'il demandait en fait des « exemples d'usage », et en a créé lui-même en indiquant comme source dans le champ prévu « invented ». Les étudiants ont alors compris qu'ils pouvaient eux aussi inventer des exemples, et l'ont fait. Au niveau du contenu, trois cas se sont produits :

- Certains ont créé ou adapté des phrases contenant les termes anglais en question, sans qu'il s'agisse de vrais « exemples d'usage », permettant de discriminer le sens en question. On trouve ainsi :

16070	collude	s'associer	they colluded last year
16990	telematics	télématique	It's telematics
16998	darts	fléchettes	he throws the darts
17003	potoling	spéléologie	the potoling is dangerous
17006	chiaroscuro	clair-obscur	the is a chiaroscuro effect
17026	heir	héritier	you heir to your mother

- En poussant cela à la limite, d'autres ont pris de longues phrases, et les ont utilisées comme exemples pour les mots ou termes de ces phrases. On trouve ainsi :

12956	Falsification	Falsification	Some various documents to be protected from counterfeiting and falsification like service vouchers, security label and certificates of authenticity have special features.
12957	service vouchers	Tickets de prestation	
12958	security label	Étiquettes sécurisées	
12959	certificates of authenticity	Certificats d'authenticité	
12960	anti-counterfeiting features	Éléments anti-contrefaçon	
12961	anti-falsification feature	Éléments anti-falsification	

- Beaucoup de propositions sont des exemples « honnêtes », mais dans un anglais incorrect. On trouve ainsi :

6619	carriageway	chaussée	the carriageway is destroy by the cars
7073	union	syndicat	an union for help employees
7098	pythonesque	humour absurde	this joke are very pythonesque with his very absurd humour
9183	(to) insulate	isoler	insulating materials can be very usefull in electronic

Ce « phénomène naturel » nous conduit à envisager d'étendre les fonctionnalités d'*IToldU* pour l'utiliser non seulement pour l'apprentissage du vocabulaire, mais aussi pour celui de la langue.

- On peut aussi mentionner un (très faible) pourcentage d'étudiants, qui se « défontent » en mettant n'importe quoi dans leurs exemples.

Au total, 15 % des exemples nous semblent incorrects du point de vue du contenu, et bien plus du point de vue de la langue, de la grammaire et de l'orthographe.

5. CONCLUSION ET PERSPECTIVES

La construction collaborative de ressources lexicales libres de droit se heurte à l'impossibilité d'obtenir beaucoup de petites contributions gratuites et volontaires. *IToldU* est un service web léger qui, lors de sa première année d'utilisation pour l'enseignement de l'anglais technique dans des écoles d'ingénieurs, a permis d'obtenir un peu plus de 17 000 termes anglais dans une vingtaine de domaines, accompagnés de leur traduction en français, correcte à 95 %, et d'exemples d'usage en français, corrects à 85 % environ. *IToldU* est facile à étendre à d'autres couples de langues. Il reste à trouver comment synchroniser *IToldU* avec *Papillon*, une base lexicale multilingue beaucoup plus ambitieuse (<http://www.papillon-dictionary.org>), et à quels autres contextes on pourrait l'étendre pour obtenir de la même façon d'autres types d'informations, comme les fonctions lexicosémantiques.

BIBLIOGRAPHIE

- Bellynck (V.), Boitet (C.) et Kenwright (J.), 2004 : « Resource pooling for technical English learning via lexical access » dans *Proceedings Papillon-04 seminar*, UJF, Grenoble, 30 Aug.-2 Sept. 2004, 5 pp.
- Bellynck (V.), Boitet (C.) et Kenwright (J.), 2005 : « *ITOLDU*, a Web Service to Pool Technical Lexical Terms in a Learning Environment and Contribute to Multilingual Lexical Databases » dans *Computational Linguistics and Intelligent Text Processing (Proc. CICLING-2005)*, A. Gelbukh, Od., Springer (LNCS 3406), pp. 319 – 327.
- Bellynck (V.), 2002 : « Bases lexicales multilingues et objets pédagogiques interactifs : Sensillon pour Papillon » dans *Proceedings of Papillon 2002 Seminar*, NII, 13 pp., Tokyo, July 2002.
- Fafiotte (G.), Boitet (C.), Seligman (M.) et Zong (C. Q.), 2004 : « Collecting Bilingual Dialogues using a Web-Based Platform for the Study of Interpretation » dans *Proc. LREC-04 (Language Resources and Evaluation Conference)*, Lisbonne, 24-28 May 2004, 9 pp.
- Mangeot-Lerebours (M.), 2001 : *Environnements centralisés et distribués pour lexicographes et lexicologues en contexte multilingue*, Ph D in computer science, Université Joseph Fourier-Grenoble I, 280 pp., Grenoble, France.
- Mangeot-Lerebours (M.), Sérasset (G.) et Lafourcade (M.), 2003 : « Construction collaborative d'une base lexicale multilingue, le projet *Papillon* » dans *TAL*, 44/2, pp. 151-176.
- Murata (T.), Kitamura (M.), Fukui (T.) et Sukehiro (T.), 2003 : « Implementation of Collaborative Translation Environment Yakushite Net », dans *Proceedings of MT Summit VIII*, New Orleans.
- Tokuda (N.) et Chen (L.), 2001 : *An Online Tutoring System for Language Translation*, *IEEE Transactions on Multimedia*, vol. VIII, n° 3, pp. 46-55, July-September 2001.

LE CHAMP SÉMANTIQUE DU MOT *MARABOUT* EN FRANÇAIS DU SÉNÉGAL

Modou Ndiaye

département de lettres modernes, faculté des lettres

Université Cheikh Anta Diop

Dakar, Sénégal

mondiaye@ucad.sn

Résumé : *En français central, le mot marabout a surtout le sens « d'homme possédant des pouvoirs magiques de devin ou de guérisseur ». En français du Sénégal, ce sens coexiste avec deux autres non moins productifs. Il peut, en effet, être l'équivalent du mot cheikh emprunté à l'arabe et désignant un chef religieux musulman, un guide spirituel. Il peut également servir à désigner le maître de Coran, c'est-à-dire un musulman qui enseigne le Coran aux enfants. Cette communication essaie de montrer que le mot « marabout » n'a pas la même grille de commutations et de combinaisons selon qu'il a le sens 1, 2, ou 3. Elle analyse les relations de commutation sur l'axe paradigmatique en examinant précisément les rapports analogiques de synonymie ou d'antonymie, et ensuite étudie les possibilités combinatoires du mot sur l'axe syntagmatique à partir d'un double point de vue : d'abord syntaxique en étudiant la structure du syntagme nominal dans lequel est pris le mot, ensuite morphologique en considérant les relations de dérivation qu'il permet.*

Mots-clés : *marabout, charlatan, devin, maître de Coran, guide religieux, champ sémantique*

1. INTRODUCTION

Le mot *marabout* est sûrement l'une de ces particularités lexicales d'Afrique qui envahissent de plus en plus le français central du fait de l'accroissement, dans les communautés d'immigrés africains vivant en France, du nombre d'individus revendiquant à titre corporatif le statut de *marabout* et favorisant du même coup les usages linguistiques qui y sont liés.

Le mot est répertorié dans le premier *Inventaire des particularités lexicales du français en Afrique noire* (IFA) comme étant d'usage dans plusieurs pays africains : Bénin, Burkina Faso, Côte d'Ivoire, Mali, Niger, Sénégal, Tchad, Togo.

Le sens qui semble prospérer le plus dans ce français central est celui, signalé par ce premier inventaire, « d'homme possédant des pouvoirs magiques de devin ou de guérisseur » (IFA, 1983).

Cependant, à côté de ce sens 1, coexistent deux autres sens non moins productifs en français d'Afrique ou tout au moins dans celui du Sénégal dont l'un (sens 2) fait du

mot l'équivalent du terme *cheikh* emprunté à l'arabe et désignant un chef religieux musulman, un guide spirituel, et l'autre (sens 3) renvoie au *maître de Coran*, c'est-à-dire à un musulman qui enseigne le Coran aux enfants.

L'objet de cette communication est de montrer que le mot *marabout* n'a pas la même grille de commutations et de combinaisons selon qu'il a le sens 1, 2, ou 3. Nous analyserons les relations de commutation sur l'axe paradigmatique en examinant précisément les rapports analogiques de synonymie ou d'antonymie, ensuite nous étudierons les possibilités combinatoires du mot sur l'axe syntagmatique à partir d'un double point de vue : d'abord syntaxique, en étudiant la structure du syntagme nominal dans lequel est pris le mot, ensuite morphologique, en considérant les relations de dérivation qu'il permet.

2. MARABOUT : SENS 1

On a l'illustration de ce sens 1 dans les exemples suivants :

[1] « Thierno Baldé, qui avait fait la connaissance de Abdoulaye Sow par l'entremise d'Amadou Baldé, s'est fait consulter par le marabout qui lui promet de le rendre riche dans un délai court » (*Le Soleil* 06/02/95).

[2] « Que c'est drôle ces histoires de *khons* et de *marabouts* dans le monde du football sénégalais » (*Le Politicien* 01/10/93).

[3] « De nombreux employés de la langue et même certains grands patrons avaient payé des féticheurs et des marabouts pour lui jeter de mauvais sort » (Akpo (A. J.) 1996 : 93).

Dans le premier exemple, le terme *marabout* désigne un homme pourvu d'un pouvoir magique de divination qui lui permet de lire le destin de ses semblables. En effet, la consultation dont il s'agit dans cet énoncé n'a rien de médical ; il s'agit d'une voyance. Cet homme a aussi le pouvoir d'agir sur le destin soit positivement – en rendant riche son client (exemple 1), ou en lui donnant la victoire (exemple 2), ou négativement – par exemple, en jetant un mauvais sort (exemple 3).

Dans ce sens 1, le terme *marabout* ne renvoie pas forcément à un musulman, ni même à un Africain, comme on est tenté de le croire. Les deux exemples suivants en attestent :

[4] « Des sources bien informées nous affirment que nos gouvernants avaient confié le destin du CFA à des marabouts recrutés un peu partout dans le monde. On parle même de féticheurs chinois, japonais et hindous » (*Sopi* 18/01/94).

[5] « Il ne reste donc plus à Ousmane Paye qu'à aller chercher un "marabout" dans le bois sacré pour que Paye soit maintenu au poste » (*Le Témoin* 14/02/96).

Ainsi, des termes génériques comme « musulman » ou « africain » ne conviendraient pas pour introduire une définition dictionnaire du mot *marabout* quand il a ce sens 1.

Dans cette acception, *marabout* forme une série paradigmaticque avec des lexies comme « féticheur, fétichiste, charlatan, guérisseur, sorcier, devin, serigne » qui en sont des synonymes. Le dernier terme¹ de la série, qui est emprunté au wolof, est en fait le seul à être strictement équivalent à *marabout*.

[6] « Il était vingt-deux heures zéro quand Léontine se présenta devant un charlatan. Après les salutations le *serigne* demanda à sa visiteuse l'objet de la promenade nocturne » (Faye (L.) 1992 : 89).

[7] « *Serigne* ! Je veux que vous voyiez pour moi l'avenir - L'avenir ? Dit le vieux *serigne* étonné » (Gaye (M.) 1995 : 64).

Par rapport aux autres lexies de la série, on a d'abord une relation d'hyponymie. Le sens de *marabout* inclut celui de ces lexies. On peut observer, par exemple, cette relation d'hyponymie dans l'énoncé (4) entre *marabout* et « féticheurs ». Cette relation pourrait être exploitée d'un point de vue lexicographique. En effet, dans la définition de ces différentes lexies, on pourrait utiliser *marabout* comme terme générique.

- Féticheur : *marabout* qui recourt aux fétiches...
- Guérisseur : *marabout* utilisant des plantes pour guérir des maladies.
- Devin : *marabout* pratiquant la divination à l'aide de *cauris* ou d'autres moyens surnaturels.
- etc.

En d'autres termes, chacun de ces mots contient le sème *marabout* comme archisé-mème. Ainsi, la relation de synonymie qu'il peut y avoir entre le terme *marabout* et ces différentes lexies est une réduction par le contexte de la relation d'hyponymie. Cette synonymie peut être établie lorsque le contexte d'emploi apporte suffisamment d'information spécifiant la modalité d'intervention ou une caractéristique précise du marabout. Ainsi, dans l'énoncé (5) par exemple, il est possible de substituer « féticheur » à *marabout* à partir de l'indication apportée par la séquence « dans le bois sacré ».

[5'] « Il ne lui reste donc plus à Ousmane Paye qu'à aller chercher un féticheur dans le bois sacré pour que Paye soit maintenu au poste. »

Cette relation de spécification fait que, plutôt que de se substituer au terme *marabout*, ces lexies peuvent simplement lui être adjointes dans une relation syntagmatique. Cela permet d'avoir les mots composés : « *marabout-féticheur*, *marabout-fétichiste*, *marabout-guérisseur*, *marabout-charlatan*, *marabout-devin*, *marabout-sorcier* ».

[8] « Elle a toujours ignoré que le bonheur n'a jamais été donné ni par l'argent, ni par les marabouts-féticheurs, mais bien, par l'économie d'une vie, dans le tronçon commun d'une vie commune » (Diop (L. S.) 1994 : 64).

¹ On retrouvera ce terme comme synonyme de « marabout » dans les deux autres sens étudiés ici.

[9] « Il est le père de Baye Guéwel griot-animateur de meeting [...] et de Djimbéra, marabout-fétichiste pour politiciens en manque de popularité » [...] (*Le Soleil* 18/05/1993).

[10] « Au potard elle se donnait enfin après qu'elle eut épuisé pour son ami tous les produits de la pharmacopée moderne [...], comme ceux de la si riche pharmacopée africaine. Et avec eux toutes sortes de pratiques, amulettes, lotions ou potions magiques des féticheurs, marabouts-guérisseurs et autres faiseurs de miracles. » (Bouso (L. M.) s. d. : 97)

[11] « Mais comme on dit, il n'y a que le premier pas qui compte. Après, tout devient facile, du moins apparemment, et cela a dû se passer ainsi pour Seynabou Diop qui, sept mois durant, a détourné de l'argent au seul profit du marabout-charlatan. » (*Le Soleil* 16/01/1995)

Dans cette série des réalisations par composition que permet le terme « marabout », on pourrait citer, également, la formation *marabout-tieddo*.

[12] « Ainsi Ndiadiane, descendant des conquérants musulmans, s'identifie à la puissance de la religion islamique qui, en Afrique noire, a élargi les limites du sacré. Le guérisseur magicien traditionnel devient un *marabout-tieddo* qui est à différencier des grands mystiques du XIX^e siècle. » (Kesteloot 1989 : 189)

Cette formation se distingue des précédentes par le fait qu'on n'a pas une relation de synonymie ni d'hyponymie entre *marabout* et le terme *tieddo*. Autrement dit, ce dernier ne pourrait se substituer au premier et occuper seul la position syntaxique. La relation entre les deux termes à l'intérieur de cette formation se limite à l'axe syntagmatique. Cette restriction est due au sens même du terme *tieddo* (emprunté au wolof) qui signifie « païen, animiste, adepte de la religion traditionnelle ». Ce sens ne contient pas, en effet, le sème *marabout* relevé dans les termes précédents.

Une analyse componentielle sommaire de la série lexicale donnerait le tableau suivant :

	Marabout	Recourt aux fétiches	Fait de la divination	Traite des maladies	Exploite la crédulité des gens
Féticheur	+	+	+	+/-	-
Guérisseur	+	+/-	+/-	+	-
Serigne	+	+/-	+/-	+/-	+/-
Devin	+	+/-	+	-	-
Fétichiste	+	+	+	+/-	-
Charlatan	+	+/-	+/-	+/-	+
Tieddo	-	+	-	-	-

Par ailleurs, le terme *marabout* donne lieu à des possibilités d'affixation qu'il ne permet que lorsqu'il a ce sens 1. En effet, on relève les dérivés par suffixation suivants :

– Marabouter : recourir à des moyens magiques ou surnaturels pour nuire à quelqu'un (ou s'attirer la chance).

[13] « Cependant les débuts n'ont pas été faciles. Il voulait s'amuser et partir. Alors [...] j'ai inventé une grossesse. Il a remué ciel et terre mais mon cinéma a fini pas le faire tomber. [...] je crois l'avoir marabouté » (*République* 20/07/1993).

– Marabouté : qui est victime d'une pratique magique maléfique.

[14] « Pour soigner d'immédiateté un ensorcellé [sic], un envoûté, un marabouté. [...] souffler dans de vache ou de brebis [...] »

– Maraboutage : action de marabouter.

[15] « Toujours est-il que les informations sont imprécises et vagues sur le mode de constitution de la fortune de telle ou telle personnalité : elles vont dans l'opinion du Sénégalais moyen du simple "maraboutage", à la pratique des prête-nom [sic], en passant par le trafic de diamants, la sorcellerie [...] » (Ndiaye (M.) 1996 : 214).

[16] « Au début, quand elle a senti ses urines sortir de son vagin, elle avait cru à un "maraboutage" ou une sorcellerie » (*Le Soleil* 16/02/1995).

– Marabouteur : personne recourant aux services d'un *marabout* pour nuire à quelqu'un. Marabout.

[17] « Parallèlement se dessinent chez les "marabouteurs" (terme qu'il faut préférer à *marabouts*, qui a une autre signification) deux tendances » (Ndione (E. S.) 1993 : 198).

– Maraboutisme : le fait social que constitue l'existence des *marabouts* ou la fonction qu'ils exercent.

[18] « Le maraboutisme, qui est une science appliquée autant que la médecine et la chirurgie... » (*Le Politicien*, 25/03/74 : 74)

– Démarabouter : combattre par des moyens magiques un envoûtement ou un mauvais sort jeté par un *marabout*.

On rencontre également l'adjectif « maraboutique » signifiant « relatif au marabout »². Cependant, cette création n'est pas limitée au sens 1. On la relève aussi au sens 2. Voici des exemples pour le sens 1.

[19] « C'est à croire qu'ils ont des rochers à la place du cœur : toutes les sortes de décoctions et de poudre [sic] maraboutiques, de cornes et de gris-gris [...] ont fini par les laisser de marbre » (*Sud Quotidien* 25/02/1995).

3. MARABOUT : SENS 2

Ce deuxième sens apparaît dans des exemples comme :

² IFA 1 a relevé également l'adjectif « maraboutal » à partir d'un exemple de Boilat, P. : « Après la bénédiction maraboutale, on fait une noce qui dure huit jours. » (Boilat (P.), *Esquisses sénégalaises* : 322). On ne retrouve pas ce terme dans les corpus récents.

[20] « Chaque année a lieu dans notre pays le Magal, qui rappelle le départ en exil du *marabout* Cheikh Ahmadou Bamba » (*Ami et Rémi* CE2 : 56).

[21] « L'alliance conclue entre certains *marabouts* et Senghor [...] était un des paradoxes criants qui caractérisent les rapports entre les chefs de la communauté musulmane et le pouvoir politique au Sénégal » (Mbacké (K.) 1995 : 118).

Marabouts signifie dans ces exemples « guide religieux ». Pour ce sens 2, la relation synonymique mène à deux termes : *serigne* et *khalife*.

[22] « Le serigne [le fondateur du mouridisme], il vous prenait de partout, vous enrobait, vous phagocytait sans crier gare » (Sougoufara, 1992 : 30).

[23] « Son fils aîné, devenu le nouveau khalife, a reçu la mission de perpétuer ce grand événement. »

À partir de ces deux synonymes sont dérivés, selon le même modèle, respectivement les termes *serignat* et *khalifat* qui désignent la fonction, le statut de guide religieux.

[24] « Le mouvement qui accorde la priorité à l'accumulation est donc, en même temps, tout empreint de la réévaluation critique du cadre social antérieur et traditionnel, de ses mœurs et de ses usages, y compris de l'institution du sérignat » (Ndiaye (M.) 1998 : 90).

[25] « C'est l'initiative la plus heureuse qu'il m'a été donné de recevoir durant mon *khalifat* » (*Promotion* n° 34 : 10).

Le terme *khalife* est emprunté à l'arabe où il signifie « premier guide spirituel ». Il garde, en fait, ce sens en wolof et en français. Il a donc un signifié qui ne correspond pas exactement à celui de *marabout*. En effet, ce dernier n'a pas forcément le sème « premier ». La relation de synonymie entre les deux termes n'est stable que si on a un contexte syntaxique dans lequel *marabout* est précédé du déterminant défini « le » et suivi d'un syntagme prépositionnel désignant le plus souvent une localité ou une confrérie.

[26] « Dans son sermon, le *marabout* de Ndiassane a rappelé aux fidèles leurs devoirs religieux. »

[27] « Cette grande manifestation religieuse sera présidée par le *marabout* des *tidianes*. »

Dans ces deux exemples, le terme *khalife* pourrait être employé à la place de *marabout*.

[28] « Dans son sermon, le *khalife* de Ndiassane a rappelé aux fidèles leurs devoirs religieux. »

[29] « Cette grande manifestation religieuse sera présidée par le *khalife* des *tidianes*. »

Dans ce contexte où le terme *marabout* est précédé du déterminant défini « le » et suivi d'un syntagme prépositionnel ayant le trait (+ localité) ou (+ confrérie), sa structure morphologique est parfois enrichie par le procédé de la composition par l'adjonction de l'adjectif « grand ».

[30] « Le chef du gouvernement et sa suite se sont rendus chez le grand *marabout* des mourides. »

Dans cet exemple, *grand marabout* a rigoureusement le sens de *khalife*, c'est-à-dire de « premier guide spirituel ».

Il faut souligner que le sens 1 ne permet pas une telle formation par composition. Il permet, certes, l'adjonction de l'adjectif « grand », mais ce n'est guère pour donner un mot composé. La séquence *grand marabout* aura, en ce cas le sens de « *marabout* de grande renommée, de talent ».

On relève également une possibilité de réalisation avec l'adjectif « jeune », qui renvoie au sens 2.

[31] Les jeunes *marabouts* ne sont pas les grands exégètes de l'islam, loin s'en faut, ce sont les gens versés dans le culte du mythe et petits fils [sic] des fondateurs de confréries ou de leurs illustres compagnons (Acart (A.) 1995 : 12).

En réalité, dans cet emploi, le sens de *marabout* est élargi à tous les membres mâles des familles de chefs religieux (fils, petits-fils, frères...), qui ont certes vocation à devenir des guides religieux, mais qui, de fait, n'ont pas forcément ce statut.

Parmi les possibilités combinatoires que permet le terme *marabout* quand il a le sens 2, figurent également des réalisations dans lesquelles il est suivi d'adjectifs spécifiques indiquant le plus souvent une activité professionnelle ou une appartenance confrérique.

[32] « *Marabout-agriculteur*, Serigne Sam Mbacké, homonyme du père de Khadimou Rassol Cheikh Ahamadou Bamba s'était installé à Ndoulo en 1945 sous le *Ndiguel* de Serigne Falilou Mbacké, alors khalife général des mourides » (*Le Soleil* 18/09/1992).

[33] « Le *marabout tidiane* sera accueilli, ce matin, à Tivaouane, par des milliers de fidèles. »

Alors que pour le sens 1 on comptait huit dérivés par suffixation, pour le sens 2 on n'en compte qu'un seul. C'est la forme « maraboutique », également relevée pour le sens 1. Voici des exemples pour le sens 2 :

[34] « La littérature anthropologique et sociologique a plutôt porté sur les relations entre l'État, les structures maraboutiques et l'économie arachidière » (Diop-Diouf 1990 : 23).

[35] « Le socialisme et la démocratie peuvent-ils exister au Sénégal face à l'omnipotence des pouvoirs maraboutiques ? » (*Le Soleil* 19/07/1993).

On relève également, pour ce sens 2, une relation d'antonymie qui n'était pas observable pour le sens 1. Elle oppose *marabout* au terme *talibé*³, emprunté au Wolof et signifiant « disciple, personne ayant fait acte d'allégeance à un marabout ».

³ Ce terme a une autre acception quand il est opposé au sens 3 de « marabout ».

[36] « Entre Serigne Modou Kara Nooreyni, le « *marabout* des jeunes » et son taalibé et admirateur Alioune Kassé, c'est (presque) la guerre » (*Le Témoin* 27/06-02/07/1995).

[37] « La cour était pleine de femmes, toutes voilées, devisant tranquillement devant le poste téléviseur [...], sans doute des *talibées* venues se recueillir auprès de leur guide » (*Le Témoin* 11-17/07/1995).

On peut observer la variation dans l'orthographe de cet antonyme (*taalibé*, ex. [36]) et son accord en genre (*talibées*, ex. [37]).

De ce terme, est formée par dérivation « talibité » qui indique l'ensemble des attributs qui font un talibé.

[38] « Grâce à son investissement, Darou Mouhty qui est le second pôle du mouridisme, abrite la plus grande mosquée de toutes les contrées mourides, demeure la référence des fidèles les plus imprégnés de la charia et de la "talibité" [...] » (*Le Soleil* 25/06/1997).

4. MARABOUT : SENS 3

Dans son troisième sens, le terme *marabout* désigne un musulman qui enseigne le Coran aux enfants. On a une illustration de ce sens dans les exemples suivants.

[39] « Les difficultés de survie en milieu rural poussent de plus en plus les *marabouts* à recourir à l'exode rural pour chercher dans les agglomérations urbaines des vivres et surtout un revenu monétaire » (*Le Soleil* 09/11/1993).

[40] « L'école coranique du marabout Yacouba Alpha Hanne [...] tape à l'œil du visiteur grâce à une écriture quelque peu maladroite, libellée en français et en arabe sur le mur de clôture » (*Le Matin* 16/04/1997).

Les relations paradigmatiques que permet le terme *marabout* dans cette troisième acception sont différentes des précédentes. En effet, ici, on a une relation de synonymie avec les formes composées « maître coranique » et « maître d'arabe », et une relation d'antonymie avec le terme *talibé*, compris non plus dans le sens déjà évoqué de « disciple, personne ayant fait acte d'allégeance à un guide spirituel » mais dans celui « d'élève coranique ». Ces deux relations sont observables dans les exemples suivants dont les trois premiers opposent « maître coranique » et *talibé* et les deux derniers *marabout* et *talibé*.

[41] « Peu de *talibés* apprennent effectivement le Coran, les maîtres coraniques préférant les envoyer mendier » (*Wal fadjri* 03/03/1988).

[42] « Avec notamment un programme d'assainissement des *daara* ciblés qui devront, au terme de cette première étape du projet, disposer d'eau potable, jusqu'à concurrence de 20 à 30 litres jour/enfant. Une initiative salutaire pour la plupart des maîtres coraniques et de leurs talibés » (*Le Soleil* 25/11/1992).

[43] « Comme le souligne d'ailleurs Oustaz Bâ, maître coranique à Grand Dakar, à propos des baby-foot qui lui portent un grand préjudice : “Tout le temps, je corrige durement mes *talibés* à cause des baby-foot” » (*Le Témoin* 28/11-05/12/1995).

[44] « L'exode vers les grandes villes étant devenu une stratégie de survie pour les enfants comme pour leurs *marabouts*, les *talibés* quittent leur terroir pendant la saison sèche à la recherche de revenus » (*Wal fadjri* 19/04/1994).

[45] « À Saint-Louis comme on le constate, le *taalibé* est un élément économique : le *marabout* compte essentiellement sur lui pour se “tirer d'affaire” » (*Wal fadjri* 18/05/1994).

Il faut signaler qu'habituellement, le *talibé* n'a pas pour seule occupation quotidienne l'apprentissage du coran. Il doit également se livrer à la mendicité pour sa propre subsistance et parfois celle du *marabout*. Ainsi, le sens du terme s'est progressivement élargi pour désigner « les jeunes enfants mendiants ». Cet élargissement de sens a entraîné la création de la lexie composée *talibé-mendiant*.

[46] « De la condition de *talibé-mendiant* à celle de bachelier, il y a un fossé énorme pour qui connaît la stratification sociale de notre pays [...] » (*Le Soleil* 23/08/1995).

[47] « Le ministre de la Santé publique a aussi fait remarquer que des études [...] ont permis [...] d'identifier les régions de Thiès, Saint-Louis et Kaolack comme étant les principales pourvoyeuses en *talibés-mendiants* » (*Le Soleil* 03/11/1995).

[48] « On compte près de 100 000 *talibés-mendiants* au Sénégal » (*Le Soleil* 24/04/1996).

Les possibilités combinatoires de *marabout* dans ce sens 3 sont presque inexistantes.

On ne relève qu'une seule réalisation par composition, par adjonction au terme *marabout* d'un substantif spécifiant la fonction d'enseignant. Il s'agit de *marabout-enseignant*.

[49] « Elles semblaient poursuivre un rêve évanescent, avec leur regard perdu dans l'horizon, où la cravache d'un *marabout-enseignant*, papillonnant au-dessus de la marmaille terrorisée, prenait un relief extraordinaire » (Ly Sangaré (D.) 1995 : 37).

[50] « Le projet de réhabilitation des *talibés* [...] s'oriente déjà vers une augmentation des capacités d'autofinancement des *daara*. Ceci, explique-t-on “pour améliorer de façon durable le niveau de vie des *marabouts* enseignants et enrayer leur exode en ville” » (*Le Soleil* 25/11/1992).

La dérivation par suffixation devrait mener à deux termes : « maraboutique » et « maraboutisme ». Cependant, aucune attestation de ces lexies (dans ce sens 3) n'est actuellement disponible en français du Sénégal.

5. CONCLUSION

Si en français central le sens 1 est le sens le plus connu, pour ne pas dire le seul connu, en français du Sénégal, il n'y a pas une telle prééminence, et tout laisse croire que le

sens 2, c'est-à-dire celui de « guide religieux », est le plus fréquent, vu l'importance du phénomène confrérique dans ce pays.

Cette communication vient confirmer l'idée que l'analyse lexicale ne devrait pas être séparée de l'analyse syntaxique⁴ et que les différences de sens d'un mot peuvent se décrire par les différences d'emplois : Les trois sens de *marabout* que nous venons de souligner correspondent, en effet, à trois emplois différents, c'est-à-dire à trois environnements différents. Dans l'étude de ces environnements nous n'avons pas dépassé les limites du SN. L'analyse pourrait être enrichie en prenant en compte les combinaisons avec les prédicats verbaux et arriver peut-être à dégager pour chaque emploi une série de verbes particuliers avec lesquels ce lexème pourrait se construire aussi bien comme sujet que comme complément.

BIBLIOGRAPHIE

Dubois (J.), 1971 : *Introduction à la lexicographie, le dictionnaire*, Paris, Larousse.

Équipe IFA, 1988 : *Inventaire des particularités lexicales du français en Afrique noire*, Paris, EDICEF-AUPELF (1^{re} éd. 1983)

Latin (D.), Ambroise (Q.) et Tabi-Manga (J.), 1993 : *Inventaire des usages de la francophonie : nomenclatures et méthodologies, Actes du colloque de Nice* (18-21 septembre 1991), Paris, AUPELF-UREF.

Pottier (B.), 1992 : *Sémantique générale*, Paris, PUF.

⁴ Cf Gross, (G.), 2005 : Communication aux septièmes Journées de formation du réseau des chercheurs en *Lexicologie, Terminologie, Traduction* organisées à Bruxelles (à l'ISTI) du 5 au 11 septembre 2005.

INTERROGER UN *CORPUS* PAR LE SENS : UNE APPROCHE LINGUISTIQUE

Bernard Jacquemin

Institut des sciences cognitives

CNRS, UMR5015

Bron, France

bernard.jacquemin@isc.cnrs.fr

Résumé : *Cet article présente une méthode d'analyse textuelle qui prend en compte la morphologie, la syntaxe et la sémantique lexicale de textes pour en identifier l'information, composée des mots significatifs, des relations (syntaxiques) entre ces mots et de leur sens. La méthode est capable de sélectionner des enrichissements pour réaliser une expansion des textes grâce notamment à des synonymes et formes dérivées intégrés aux documents pour en reformuler le texte sans pratiquement modifier son sens, afin de l'interroger par le sens.*

Mots-clés : *question-réponse, désambiguïsation sémantique, expansion, synonymie, dérivation, dictionnaire*

1. INTRODUCTION

La société actuelle a fait de la maîtrise de l'information un enjeu de savoir autant que de pouvoir. Cependant, face à la profusion des sources d'information, face à l'enchevêtrement ingérable des données elles-mêmes, personne n'est plus capable de fournir un accès rapide à un élément d'information précis. Les initiatives qui visent à élaborer une méthode automatique de gestion de l'information capable d'ordonner des masses de données sont dès lors bienvenues. Dans les champs de recherche liés à l'information textuelle électronique, et notamment la tâche de question-réponse, plusieurs méthodes ont vu le jour, qui permettent de confronter les données de la question avec celles contenues dans un texte. Si les données correspondent, on considère que la réponse à la question est dans le contexte immédiat de l'information commune à la question et à la bribe de texte. Diverses conférences internationales ont également vu le jour, dont l'objet est l'évaluation des systèmes proposés : TREC, CLEF, NTCIR...

Il reste toutefois que toutes ces méthodes fonctionnent sur base d'un même schéma. Il s'agit en effet d'appréhender la question, de l'analyser pour la débarrasser de tout élément perturbateur, et d'en effectuer une expansion destinée à contrer les variations de forme de l'information dans les documents interrogés. De plus, les meilleurs systèmes plaident tous pour des approches capables de gérer le caractère langagier des textes. Par exemple, Hull (1999) propose l'exploitation des résultats d'une analyse morphologique pour indexer les éléments significatifs tant dans les requêtes que dans les réponses. Ferret *et al.* (2002) présentent une certaine intégration de la syntaxe, capable de recon-

naître des entités nommées et de déterminer la nature de la réponse à la question. Le meilleur système disponible s'appuie sur des données sémantiques issues du réseau WordNet, ainsi que sur un moteur d'inférences logiques pour atteindre un niveau élevé de réponses correctes (Harabagiu *et al.* 2000).

Une consultation, même rapide, des publications liées aux campagnes d'évaluation du domaine de question-réponse (Voorhees et Buckland 2004, Peters *et al.* 2005, Kando et Ishikawa 2005) montre que les différentes approches se distinguent par leur capacité à appréhender plus ou moins bien la langue, à obtenir une meilleure analyse linguistique non seulement de la question, mais également des réponses possibles. Malgré les vertus reconnues des modules d'analyse linguistique, il est pourtant étrange de constater qu'ils n'occupent qu'une place générique dans toutes ces approches à dominante statistique, et qu'aucune recherche n'est actuellement menée pour leur accorder un statut plus central qui pourrait encore améliorer le fonctionnement des logiciels.

Cet article s'appuie sur les constats précédents pour proposer une méthode où l'analyse linguistique occupe une place prépondérante à tous les niveaux du système. Nous allons d'abord présenter les contraintes propres à une analyse morphosyntaxique et sémantique de textes et exposer les choix auxquels elles nous ont amené. Ensuite, nous présenterons brièvement les outils d'analyse utilisés ici, ainsi que les ressources lexico-sémantiques exploitées, et les adaptations nécessaires. Après cela, nous présenterons la construction d'une structure informationnelle qui fournit un accès à chaque élément d'information d'une base textuelle. Enfin, nous présenterons la méthode permettant de trouver la réponse à une question posée en français dans cette base textuelle. Finalement, nous exposerons quelques perspectives pour ce sujet de recherche, et notamment l'exploitation de cet outil en étude de *corpus* linguistiques.

2. UNE MÉTHODE LINGUISTIQUE DE STRUCTURATION TEXTUELLE

La stratégie présentée est centrée sur une analyse linguistique des énoncés. Cette démarche s'inscrit toutefois dans la tradition du domaine, puisque l'information considérée est définie dans un courant syntaxico-sémantique et lexico-sémantique. Il s'agit en effet d'identifier les unités lexicales porteuses de sens, considérées comme l'information élémentaire, les relations (syntaxiques) entre ces éléments, ainsi que le sens lexical des lexèmes en contexte. D'autre part, les variations dans la formulation d'une même information sont classiquement compensées par une expansion de l'énoncé, à travers des synonymes, hyperonymes, holonymes et autres dérivés morphologiques.

L'expansion d'énoncé pratiquée dans la tâche de question-réponse s'applique généralement à la requête. Le principe en est simple : à chaque unité lexicale significative est associée une liste de mots jugés équivalents, sous forme de synonymes, de dérivés, etc. Ces listes d'expansions sont utilisées disjonctivement au lexème original lorsque la requête est proposée à un moteur de recherche. Mais si une telle expansion permet effectivement de résoudre dans un grand nombre de cas les problèmes de formulation, seule la maîtrise du sens de l'énoncé à «*expanser* » permet de sélectionner les reformulations qui conviennent dans le contexte courant. La figure 1 distingue les expansions

correctes (A) des expansions erronées (B), susceptibles d'apporter des réponses inadéquates.

Question :

« De quel **chef** Domitien est-il le **successeur** ? »

A	général empereur ...	héritier succéder ...
B	cuisinier cheveu ...	remplaçant succédané ...

Réponse :

Second fils de Vespasien, Domitien succéda à l'empereur Titus et poursuivit la remise en ordre de l'État.

Figure 1 : Exemple d'expansion d'énoncé : le problème du sens

Une sélection du sens des unités lexicales de la requête est donc nécessaire pour que l'expansion puisse être effectuée en fonction du sens original, afin de limiter le bruit. Cependant, on sait depuis Weaver (1949) toute l'importance que prend le contexte – particulièrement syntaxique (Reifler 1955) – dans le choix du sens lexical d'un mot dans une phrase (désambiguïsation sémantique lexicale). De plus, les questions que l'on pose au système sont généralement plus courtes que des phrases rédigées dans un document. Le contexte y est donc moins important que dans les textes interrogés, et la syntaxe particulière de l'interrogation modifie également la qualité de ce contexte. Il est donc moins efficace de traiter la sémantique lexicale d'un lexème s'il apparaît dans une requête que s'il figure dans la base textuelle elle-même.

Dès lors, une démarche qui applique prioritairement une analyse syntaxico-sémantique aux documents plutôt qu'à la requête s'est imposée. Ce choix est d'autant plus opportun qu'il s'intègre pleinement à l'indexation du contenu lexical des documents, indispensable pour la phase de recherche, et qui consiste à recenser le lexique des documents. De plus, une telle approche centrée sur les documents présente comme avantage la distinction chronologique de l'analyse–expansion des énoncés, et la phase d'interrogation. Les traitements les plus lourds sont donc appliqués préalablement, et l'interrogation de la structure de l'information est presque instantanée.

Le système comporte dès lors deux niveaux de fonctionnement. Le premier consiste à analyser les documents d'un point de vue morphologique, syntaxique et sémantique, puis à leur appliquer une expansion à l'aide d'informations provenant de ressources lexico-sémantiques, et à stocker les résultats dans des index constituant une structure de

l'information textuelle (Roux et Jacquemin 2002). La seconde étape a pour objet l'interrogation de cette structure à l'aide de questions ordinaires.

3. OUTILS D'ANALYSE TEXTUELLE

Divers outils d'analyse interviennent dans l'identification des éléments d'information présents dans les documents. Il s'agit d'identifier les éléments eux-mêmes, c'est-à-dire les mots significatifs, au travers d'une analyse morphologique ; ensuite, les relations entre ces mots grâce à l'analyse syntaxique ; enfin, la désambiguïsation sémantique permet de connaître la signification des mots dans leur contexte d'apparition. Voici une rapide description de ces outils.

3.1. L'analyseur morphologique

L'analyseur morphologique NTM (*Normalizer, Tokenizer, Morphological analyzer*) que nous avons utilisé est un système de transducteurs à états finis développé par Aït-Mokhtar (1998) au Centre de recherche européen de Xerox (XRCE). Ce système prend en entrée n'importe quelle chaîne de caractères en français et y applique des traitements de normalisation, de segmentation s'il s'agit de plusieurs unités lexicales, et propose les différentes analyses morphologiques possibles pour chaque segment identifié. La figure 2 permet d'identifier les traitements appliqués par NTM à une phrase proposée en entrée. Il permet d'obtenir une version normalisée de chaque unité lexicale, son lemme et les informations morphologiques qui y sont associées.

Son deuxième fils [...]

Mot du texte	lemme	Analyse morphologique	Traits ajoutés
son	son	+PP3S+InvGen+SG+Poss	+SOM+AGR
son	son	+Masc+SG+Noun+	

PP3S	Pronom personnel 3ème sg	Masc Masculin
InvGen	Invariable en genre	SG Singulier
SG	Singulier	Noun Nom
Poss	Possessif	SOM Relatif au corps AGR Agriculture

Figure 2 : Exemple d'analyse morphologique par NTM

Cet analyseur présente également une qualité qui a pu être exploitée avec succès. En effet, sa constitution sous forme de lexiques – transducteurs permet d'ajouter aisément certaines informations lexicales utiles pour les traitements ultérieurs. Ainsi, on peut voir dans l'exemple que nous avons ajouté aux lexiques existants des informations sémantiques extraites d'un dictionnaire, qui seront utilisées ultérieurement lors de la phase de désambiguïsation. Dans l'intervalle, cette information subsiste, attachée aux unités lexicales, mais elle reste virtuelle dans la mesure où elle n'intervient ni dans la désambiguïsation catégorielle, ni dans l'analyse syntaxique.

3.2. L'analyseur syntaxique

L'analyseur syntaxique XIP (*Xerox incremental parser*) (Roux 1999) est un l'outil d'analyse syntaxique de XRCE, basé sur des grammaires de réécritures incrémentales. Il permet d'effectuer le cas échéant une désambiguïstation catégorielle d'énoncés étiquetés morphologiquement mais non désambiguïsés. Il propose surtout une analyse syntaxique de surface robuste de ces énoncés sous forme de dépendances entre des nœuds représentés par les unités lexicales qui sont les têtes des syntagmes minimaux (*chunks*) concernés. Une représentation en arbre de chaque phrase, ainsi qu'un découpage en syntagmes minimaux sont également proposés, mais ils ne sont pas utilisés ici.

Énoncé : « Il reconstruisit Rome ruinée par les incendies. »

Extraction des dépendances :

SUBJ (reconstruisit, Il)	2e argument sujet du 1er argument
SUBJ (ruinée, incendies)	
VMOD [INDIR] (ruinée, par, incendies)	3e argument compl. agent 1er argument
VARG [DIR] (reconstruisit, Rome)	2e argument COD du 1er argument
NMOD [ADJ] (Rome, ruinée)	2e argument épithète du 1er argument

Figure 3 : Exemple d'analyse syntaxique par XIP

La figure 3 permet d'évaluer les possibilités de XIP et d'illustrer son mode de représentation des relations syntaxiques par dépendances. On distingue certains traits, exclusivement syntaxique ici, qui portent sur les dépendances (DIR et INDIR respectivement sur les dépendances VMOD et VARG, ainsi que ADJ sur NMOD). Des traits peuvent également être associés aux arguments. XIP applique des règles contextuelles qui permettent d'évaluer des nœuds et des traits portant sur des nœuds appartenant à un même contexte pour construire des syntagmes minimaux et des dépendances. Ce mode de fonctionnement est très intéressant : il permet de travailler sur des indications lexicomorphologiques pour la désambiguïstation catégorielle et des données lexicosyntaxiques pour la construction des dépendances syntaxiques. Il n'y a donc pas d'obstacle à son utilisation dans une perspective lexico-sémantique.

3.3. Le désambiguïseur sémantique

Le système de désambiguïstation sémantique présenté ici est une évolution de la méthode de Brun *et al.* (2001), qui reposait sur l'exploitation de l'analyse syntaxico-sémantique d'un dictionnaire utilisé comme *corpus* sémantiquement étiqueté. Le présent système (Jacquemin *et al.* 2002) exploite l'information du *Dictionnaire des verbes français* de Dubois et Dubois-Charlier (1997) et de son complément des autres catégories grammaticales (désormais *Dubois*). Ces dictionnaires répartissent l'information fournie non par unité lexicale, mais par sens de chaque unité lexicale. De la sorte, cha-

que information fournie par le dictionnaire est discriminante pour le sens concerné d'un mot donné.

Le fonctionnement du désambiguïseur sémantique se fait en deux temps : d'abord l'analyse du dictionnaire avec création de règles conditionnelles de désambiguïsement, basées sur un schéma syntaxique, et ensuite l'application de ces règles à des mots en contexte, sur base des contextes syntaxico-sémantiques fournis d'abord par les étiquettes sémantiques ajoutées à NTM, ensuite par les dépendances issues de l'analyse syntaxique.

La nature de l'information lexico-sémantique régit le type de règles envisageables. Dans le cas du Dubois, l'information peut être diverse et se présenter sous forme purement syntaxique (p. ex. « Je bois » vs « Je bois de l'eau » : transitivité), syntaxico-sémantique (p. ex. « embrasser quelque chose » vs « embrasser quelqu'un » : sous-catégorisation du complément direct), lexico-syntaxique avec le schéma syntaxique des exemples considéré comme typique du sens concerné (« le général remporte la victoire » implique la dépendance VARG [DIR] (remporter, victoire) : *victoire*, complément direct de *remporter*) ou sémantico-syntaxique (généralisation du schéma syntaxique d'un exemple par les traits sémantiques d'une unité lexicale : VARG [DIR] (remporter, [MIL]) : trait MIL, militaire, de *victoire* dans le schéma).

Comme les règles de désambiguïsement doivent répondre au contexte syntaxique, selon Reifler (1955), et qu'elles sont conditionnelles – puisque la conformité d'un contexte à une information issue d'un dictionnaire implique la sélection du sens correspondant – elles répondent aux caractéristiques des règles des grammaires XIP. C'est donc à cette syntaxe que les règles de désambiguïsement doivent correspondre, ce qui permet d'utiliser XIP comme moteur sémantique. Les résultats d'une désambiguïsement sémantique apparaîtront donc comme des dépendances extraites par XIP ou comme des traits sur des dépendances ou des nœuds de XIP. La figure 4 permet de comprendre le mode de construction de règles de désambiguïsement sémantique à partir de l'information contenue dans le dictionnaire. Les données syntaxiques ou lexicales sont formalisées sous la forme d'une dépendance XIP et les données sémantiques sous la forme de traits sur les nœuds.

L'application des règles se fait à travers l'analyse des énoncés par XIP. La mise en correspondance de l'analyse syntaxique avec le schéma lexico-sémantico-syntaxique d'une règle de désambiguïsement implique l'application de la règle et la sélection du sens correspondant. Le sens sélectionné est associé sous la forme d'un trait à l'unité lexicale considérée. Les informations sémantiques ajoutées par NTM servent à l'application des règles impliquant des indications sémantiques.

Exemple extrait du Dubois pour « *remporter* » au sens 03 *gagner* :

« Le général remporte la victoire ».

Dépendances extraites de l'exemple :

SUBJ(*remporter*, général)

VARG[DIR](*remporter*, victoire)

- Construction d'une règle lexico-syntaxique de désambiguïsation :

remporter : VARG[DIR](*remporter*, victoire)

⇒ *remporter* 03 « *gagner* »

→ apparition de victoire comme complément direct de *remporter* implique le sens 03 « *gagner* »

- Construction de la règle sémantico-syntaxique correspondante :

victoire → trait sémantique : MIL (militaire)

remporter : VARG[DIR](*remporter*, MIL)

⇒ *remporter* 03 « *gagner* »

→ apparition d'un mot comportant le trait MIL (militaire) comme complément direct de *remporter* implique le sens 03 « *gagner* »

Figure 4 : Exemple de construction des règles de désambiguïsation

4. ADAPTATION DES RESSOURCES LEXICALES

Les données lexicales sont primordiales non seulement dans la perspective de l'analyse sémantique, où l'information syntaxico-sémantique distribuée par sens des entrées est prépondérante, mais aussi dans une optique d'expansion d'énoncés. En effet, cette expansion est effectuée par remplacement d'unités lexicales originales par d'autres avec un minimum de modifications de sens. Deux types de modifications lexicales sont réalisés : la synonymie et la dérivation morphologique. Dans une certaine mesure, le dictionnaire Dubois est à même de fournir les indications permettant de procéder à ces transformations.

En effet, un des champs informationnels de ce dictionnaire de référence fournit des synonymes, tandis qu'un autre procure les indications de dérivation. Toutefois, les synonymes sont invariablement au nombre de deux, ce qui est généralement insuffisant pour couvrir l'ensemble des transformations synonymiques possibles. D'autre part, les indications de dérivations, basées sur une racine et des suffixes, sont difficilement in-

terprétables directement par un système automatique. Dès lors, d'autres ressources et outils doivent être exploités pour combler les lacunes du Dubois.

4.1. Adjonction et répartition de synonymes

Pour ajouter une information synonymique au Dubois, nous avons utilisé trois ressources lexico-sémantiques : *EuroWordNet* (Vossen 1998, Catherin, 1999), le *Dictionnaire des synonymes* de Bailly (1947), et un dictionnaire multilingue de Memodata. Tous fournissent des synonymes, mais leur éventuelle répartition par sens ne correspond pas à celle du Dubois. Il a donc fallu les redistribuer. Nous avons élaboré une méthode qui le fait automatiquement (Jacquemin 2004b).

Cette procédure établit pour chaque entrée de chaque nouveau dictionnaire la liste des synonymes sans distinguer les sens différents de cette entrée. Ensuite, à chaque synonyme proposé, elle associe toutes les étiquettes sémantiques qui lui sont attachées dans le dictionnaire Dubois. Puis, pour l'entrée considérée, chaque sens du Dubois est considéré successivement : lorsqu'une des étiquettes sémantiques du synonyme proposé est identique à celle du sens courant de l'entrée, il est considéré comme un synonyme valable pour ce sens et ajouté au champ de synonymie du Dubois. La même opération est effectuée pour chaque entrée de chaque dictionnaire de synonymes, puis les doublons sont éliminés. La figure 5 illustre la procédure suivie.

ravir	sens n°2, « voler » étiquette sémantique SOC	
Synonymes proposés :	enlever	étiquette sémantique SOC / LOC / TEX... PSY / OCC
	charmer	

	⇒ synonyme ajouté à ravir(02) : enlever	

Figure 5 : Répartition des synonymes par sens du mot original

4.2. Génération de dérivés

D'autre part, l'information du Dubois ne permet pas de générer automatiquement les formes dérivées à partir d'une vedette du dictionnaire. Par contre, cette information est suffisante pour valider ou non une proposition de dérivation. L'outil de dérivation morphologique que propose Gaussier (1999), qui permet de générer, pour un mot proposé, un très grand nombre de candidats dérivés attestés dans le lexique, peut dès lors se révéler utile. Les données de dérivation indiquées dans les champs correspondants du Dubois permettent ensuite, par identification du suffixe et de certaines caractéristiques de la racine, de ne conserver pour chaque sens que les dérivés prescrits par le dictionnaire.

5. CONSTRUCTION DE LA STRUCTURE INFORMATIONNELLE

Comme on l'a vu plus haut, la structure de l'information est constituée d'index comprenant l'ensemble des données contenues dans les dictionnaires, et permettant d'avoir accès directement à la bribe de texte intéressante dans la base documentaire. Cette structure est constituée d'abord du résultat de l'analyse des textes, c'est-à-dire des unités lexicales identifiées lors de l'analyse morphologique, ainsi que des traits morphologiques qui y sont associés, puis des relations syntaxiques entre ces lexèmes, ainsi que des traits syntaxiques associés soit aux dépendances, soit aux unités lexicales, et enfin des traits sémantiques dénotés lors de la désambiguïsation sémantique (numéro de sens, indications sémantiques propres à ce sens dans le Dubois), qui portent uniquement sur les unités lexicales. La figure 6 montre comment la structure informationnelle est construite à partir d'un énoncé et de son analyse : les traits sont représentés entre crochets, les dépendances en majuscules et les unités lexicales en minuscules.

« ...Domitien succéda à l'empereur Titus... »

```
SUBJ (succéda [sn=1], Domitien [proper])
VARG [INDIR] (succéda [sn=1, ], à, empereur [humain, ])
NN (empereur, Titus [proper])
```

Figure 6 : Construction du « squelette » de la structure informationnelle

Dans un deuxième temps, la structure informationnelle est enrichie par l'expansion des énoncés. Les synonymes sont ajoutés disjonctivement aux dépendances dans lesquelles apparaissent les unités lexicales originales. Par contre, les formes dérivées ne peuvent être placées de même dans la structure informationnelle, car elles appartiennent souvent à une catégorie grammaticale différente de celle de leur lexème original, et n'obéissent pas à un schéma syntaxique similaire. Pour conserver une signification proche de celle de l'énoncé original, il s'agit donc de reformuler l'énoncé pour qu'il intègre la forme dérivée. Pour cela, nous avons étudié le processus de dérivation : pour chaque catégorie grammaticale originale, pour chaque catégorie grammaticale dérivée, pour chaque type suffixal de dérivation, nous avons sélectionné au hasard trois exemples de dérivation dans le Dubois, et observé de quelle manière on peut remplacer l'original par le dérivé dans vingt contextes réels obtenus sur le Web. À partir de là se sont dégagés des patrons de dérivation qui permettent d'identifier le schéma syntaxique d'apparition de l'original et d'inférer un schéma syntaxique de dérivation.

La figure 7 permet de comprendre de quelle manière les différents enrichissements sont ajoutés à la structure originale pour constituer des expansions des énoncés originaux, que ce soit par synonymie ou dérivation.

« ...Domitien succéda à l'empereur Titus... »

Résultats avant expansion :

```

SUBJ (succéda [sn=1] , Domitien [proper] )
VARG [INDIR] (succéda [sn=1] , , à , empereur [humain] , )
NN (empereur , Titus [proper] )

```

Structure de l'énoncé avec expansion :

Synonymie

<pre> SUBJ (succéda/remplacer , Domitien) VARG [INDIR] (succéda , à , empereur) VARG [DIR] (remplacer , empereur/chef/[...]) NN (empereur/chef/souverain/dots , Titus) </pre>

Dérivation

<pre> NMOD [INDIR] (successeur , de , empereur/chef/souverain/...) NMOD (Domitien , successeur) </pre>
--

Figure 7 : Construction de la structure informationnelle avec expansion

6. INTERROGATION DE LA STRUCTURE

Dans le cadre de l'application de question-réponse, le système doit apporter une réponse à une question en langue naturelle (Jacquemin 2004a). L'information contenue dans la question doit donc être convertie dans un format similaire à celui de la structure informationnelle, c'est-à-dire dans une structure locale similaire. Toutefois, comme le contexte d'une question est insuffisant pour effectuer une analyse sémantique congrue, cette structure locale est légère, c'est-à-dire qu'elle est limitée aux analyses morphologique et syntaxique, excluant donc la désambiguïsation sémantique et la phase d'expansion.

Certaines particularités doivent pourtant être signalées dans la conception de cette structure légère de la question. En effet, une grammaire particulière est mise en œuvre dans l'analyse de la question, qui permet deux adaptations de la structure. La première réside dans une dépendance FOCUS, qui permet de caractériser l'objet de la question, et donc la réponse attendue. Il s'agit d'une dépendance de marquage, qui identifie l'unité lexicale la plus significative de l'interrogation, c'est-à-dire la tête du groupe nominal lorsque l'interrogatif est un adjectif (« Qui est le beau-père de Galère ? » FOCUS (beau-père [PAR])) ou l'interrogatif lui-même si c'est un pronom (« Qui combattit les Parthes ? » FOCUS (qui [humain])). Elle permet d'identifier les traits sémantiques (entre crochets) de l'objet de la question, et donc d'identifier la réponse dans les documents lorsque les autres éléments de la question se trouvent dans un énoncé de la base textuelle. Cette dépendance n'existant ni dans les documents, ni dans la structure informationnelle, elle devra ensuite être transmise comme un trait à l'intérieur de la

structure légère. Dans un même temps, elle devra être supprimée en tant que dépendance dans cette structure légère.

La seconde adaptation de la structure locale réside dans la suppression de toutes les informations ne relevant que du caractère interrogatif de cette question. Ainsi, l'interrogatif sera supprimé pour ne conserver, au sein des dépendances qui le contiennent, que les traits sémantiques qui lui sont propres et, le cas échéant, le trait FOCUS. Les dépendances purement fonctionnelles disparaissent également (dues au fonctionnement interne de XIP ou mettant en œuvre des mots-outils, des auxiliaires ou semi-auxiliaires, etc.), car elles ne sont pas porteuses d'information pertinente dans le cadre de cette application. La dépendance FOCUS transmet son information sous forme de trait au lexème qui en constitue l'argument, et disparaît de même.

« De quel chef Domitien est-il le successeur ? »

Structure légère de la question :

SUBJ (est, Domitien)
FOCUS (chef)

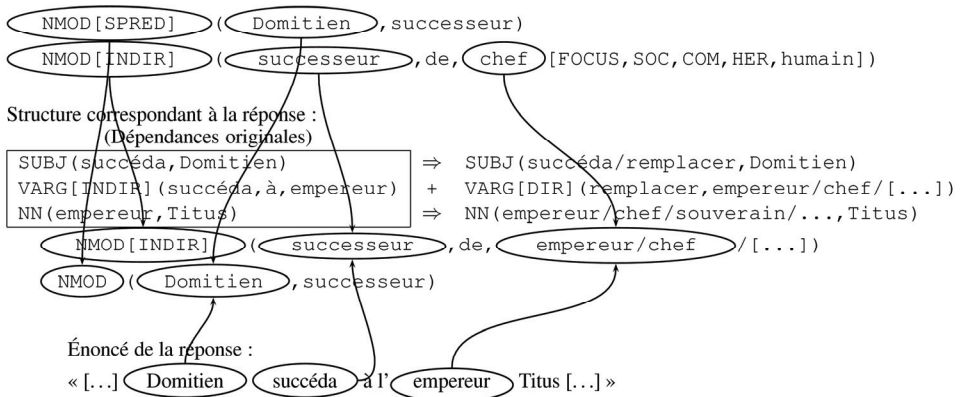


Figure 8 : Exemple d'interrogation de la structure informationnelle avec son expansion

La recherche d'une réponse revient donc à mettre en correspondance la structure légère de la question, débarrassée de l'information propre à une interrogation, et des bribes de texte à travers leur structure informationnelle. Lorsqu'une information concordante à la structure légère est trouvée au sein de la même phrase dans la structure de l'information, cette phrase est considérée comme une réponse pertinente à la question. Bien entendu, la réponse est considérée comme plus pertinente si une plus grande partie de l'information qui concorde est originale dans le texte, et moins pertinente à mesure que ces éléments concordants sont issus d'une expansion. La figure 8 illustre bien

la mise en concordance d'une question avec sa réponse au travers de deux structures, l'une légère et purifiée, l'autre complète et enrichie d'expansions.

7. CONCLUSION

Nous avons présenté un système généraliste d'interrogation d'une base documentaire textuelle en langue naturelle. Ce système s'appuie sur des bases théoriques et sur des constatations pratiques pour proposer une méthode originale de structuration de l'information dans une base textuelle avec expansion des documents plutôt que des requêtes. L'ensemble des analyses et des enrichissements ont été effectués par des analyseurs linguistiques et leurs analyses réalisées suivant des indices contextuels et symbolistes à l'aide de grammaires décrivant la langue.

Une évaluation de ce système n'a pu être présentée ici faute de place. On peut en trouver le détail dans Jacquemin (2003). Elle montre la validité de la méthode – qui soutient la comparaison avec les meilleurs systèmes de sa catégorie dans la conférence TREC9 –, ainsi que certaines faiblesses, essentiellement liées à l'absence de résolution d'anaphores ou de hiérarchie sémantique. D'autre part, cette approche souffre, comme c'est habituel dans le domaine, de la représentation exclusivement lexicale de l'information, qui tient peu compte des mécanismes logiques. L'inférence, par exemple, n'est pas gérée actuellement, mais certaines approches statistiques du lexique sont prometteuses.

Par ailleurs, la présentation de ce système a été faite uniquement dans une optique de gestion de l'information. Cependant, il pourrait également se révéler un précieux outil d'étude de *corpus* linguistique écrit, dans la mesure où il peut être interrogé aisément et rapidement, que tous les niveaux d'information linguistique sont disponibles à tout moment et qu'ils peuvent être individualisés sans problème. Ainsi, on peut facilement mêler dans une même requête des exigences lexicales, morphologiques, syntaxiques, sémantiques, de cooccurrence, et obtenir l'ensemble des réponses pertinentes quel que soit le *corpus* désiré, puisque ce système est automatique et qu'il accepte du texte tout venant avec une robustesse remarquable. Une telle approche à dominante linguistique semble dès lors se justifier, même si des améliorations peuvent et doivent y être apportées.

BIBLIOGRAPHIE

- Aït-Mokhtar (S.), 1998 : *L'analyse présyntaxique en une seule étape*, thèse de doctorat, Université Blaise Pascal Clermont-Ferrand II.
- Bailly (R.), 1947 : *Dictionnaire des synonymes de la langue française*, Paris, Larousse.
- Brun (C.), Jacquemin (B.), Segond (Fr.), 2001 : « Exploitation de dictionnaires électroniques pour la désambiguïsation sémantique lexicale », dans *Traitement Automatique des Langues*, t. 42 (3), pp. 667 – 690.
- Catherin (L.), 1999 : *The French WordNet*, rap. tech., EuroWordNet.

- Dubois (J.), Dubois-Charlier (F.), (1997) : *Dictionnaire des verbes français*, Paris, Larousse. La première version de ce dictionnaire est électronique. Elle est accompagnée de son complément *Dictionnaire des mots*.
- Ferret (O.), Grau (B.), Hurault-Plantet (M.), Illouz (G.), Monceaux (L.), Robba (I.), Vilnat (A.), 2002 : « Recherche de la réponse fondée sur la reconnaissance du focus de la question », dans *Actes de TALN 2002*, pp. 98 – 107.
- Gaussier (É.), 1999 : « Unsupervised learning of derivational morphology from inflectional lexicons », dans *ACL'99 Workshop Proceedings on Unsupervised Learning in Natural Language Processing*, ACL'99, College Park, Maryland, USA, pp. 24 – 30.
- Harabagiu (S.), Moldovan (D.), Pscu (M.), Mihalcea (R.), Surdeanu (M.), Bunescu (R.), Gîrju (R.), Rus (V.), Morarescu (P.), 2000 : « FALCON : Boosting Knowledge for Answer Engines », dans *Proceedings of Text REtrieval Conference (TREC-9)*.
- Hull (D. A.), 1999 : « Xerox TREC-8 Question Answering Track Report », dans *Proceedings of The Eighth Text Retrieval Conference*, TREC-8, pp. 743 – 752.
- Jacquemin (B.), 2003 : *Construction et Interrogation de la structure informationnelle d'une base documentaire en français*, thèse de doctorat, Université Paris III-Sorbonne nouvelle, Paris. <http://halshs.ccsd.cnrs.fr/halshs-00003957>.
- Jacquemin (B.), 2004a : « Analyse et expansion des textes en question-réponse », dans Purnelle (G.), Fairon (C.), Dister (A.), édit., *Le Poids des mots. Actes des 7^{es} journées internationales d'analyse statistique des données textuelles*, JADT04, Presses universitaires de Louvain, Louvain-la-Neuve, t. II, pp. 633 – 641.
- Jacquemin (B.), 2004b : « Dictionaries merger for text expansion in question answering », dans Zock (M.), Saint-Dizier (P.), édit., *Proceedings of COLING. Enhancing and using electronic dictionaries*, COLING04, Genève.
- Jacquemin (B.), Brun (C.), Roux (C.), 2002 : « Enriching a text by semantic disambiguation for information extraction », dans de Lopy (C.), édit., *LREC 2002 Workshop Proceedings. Using Semantics for Information Retrieval and Filtering*, CIRLEP (Centre interdisciplinaire de recherches en linguistique et psychologie cognitive), Las Palmas de Gran Canaria, Espagne.
- Kando (N.), Ishikawa (H.), éd., 2005 : *Proceedings of the Fourth NTCIR Workshop on Research in Information Access Technologies, Information Retrieval, Question Answering and Summarization*, NTCIR Workshop 4, National Institute of Informatics.
- Peters (C.), Clough (P.), Gonzalo (J.), Jones (G. J. F.), Kluck (M.), Magnini (B.), édit., 2005 : *Multilingual Information Access for Text, Speech and Images. 5th Workshop of the Cross-Language Evaluation Forum*, t. 3491, CLEF 2004, Kluwer, Bath, UK.
- Reifler (E.), « The mechanical determination of meaning », dans Locke (W. N.), Booth (A. D.), édit., 1955 : *Machine Translation of Languages : Fourteen Essays*, John Wiley & Sons, Clapham & Hall, New York, London, pp. 136 – 164.
- Roux (C.), 1999 : « Phrase-driven parser », dans *Proceedings of VEXTAL '99*, VEXTAL'99, Venezia, Italia, pp. 235 – 240.
- Roux (Cl.), Jacquemin (B.), *Storing and indexing of each level of the inner structure of a document with binary vector indexes, with the deepest level being the result of natural language processes*, Dépôt de proposition de brevet, Xerox, 2002.
- Voorhees (E. M.), Buckland (L. P.), édit., 2004 : *The Thirteenth Text REtrieval Conference Proceedings*, TREC 2004, NIST, 2004.
- Vossen (P.), édit., 1998 : *EuroWordNet : a multilingual database with lexical semantic networks*, Kluwer Academic Publishers, New York. réédition de *Computer and the Humanities*, 32 (2-3), 1998.
- Weaver (W.), 1949 : « Translation », dans Locke (W. N.), Booth (A. D.), édit., 1955 : *Machine Translation of Languages : Fourteen Essays*, John Wiley & Sons, Clapham & Hall, New York, London, pp. 15 – 23.

QUELS CORPUS ET QUELLES APPROCHES POUR UNE DESCRIPTION CONTRASTIVE DE L'EUROLECTE ?

Roger Goffin

Université libre de Bruxelles
Bruxelles, Belgique

Résumé : *La présente étude s'inscrit dans le cadre d'une réflexion d'ensemble, entamée en 1990, sur la nature de l'eurolecte. C'est le nom que nous avons donné au technolecte spécifique que s'est forgé l'Union européenne pour décrire l'ordre juridique autonome auquel elle a donné naissance.*

Il vise à dénommer les concepts en devenir et à maîtriser l'afflux d'unités lexicales nouvelles (eurolexies néologiques) d'une Europe en pleine gestation.

Nous tentons d'abord de décrire les vastes corpus de matériaux textuels sur supports traditionnels ou informatisés, tirés notamment du droit originaire (p. ex. les traités), du droit dérivé ou des politiques communes, qui entrent dans la matière de recherche.

Nous explorons par ailleurs la nature du texte (texte originaire), les apports d'une approche contrastive, les tentatives d'améliorer la lisibilité des écrits.

Mots-clés : *eurolecte, corpus, Union européenne, consolidation, eurolexie*

1. HARO SUR « L'EUROJARGON »

Le langage que charrient les écrits communautaires (Parlement, Conseil, Commission) a été la cible de toutes les attaques. Il ne cesse d'être fustigé par les hommes politiques, vilipendé par les eurodéputés eux-mêmes, la presse le pourfend avec une hargne polémique. Les médias se sont très tôt emparés de ces termes, et, en les poussant sur le devant de la scène, ont brouillé l'image qu'ils donnent d'eux-mêmes dans leur contexte européen. Les censeurs, nombreux et impitoyables, dénoncent en français l'*euroabillage*, le *brouillard linguistique européen*, le *volapuk intégré*, le *sabir habituel de l'Officine de Bruxelles*. L'allemand n'est pas à la traîne puisqu'il a forgé *Euronebel*, *Eurowelsch* ou *Eurokauderwelsch* (*Kauderwelsch* signifie charabia ou baragouin) ; en anglais, *Eurospeak*, *Eurofog*, *Eurobabble*. Marcel Thiry (1897-1977), poète et eurodéputé belge, assommé par le *parlage* linéaire et gris des réunions européennes, se met à rêver des fausses muettes que sont les interprètes, qui ont incendié par les « mobilités de leurs lèvres muettes » l'imagination du poète.

Les procureurs s'en prennent tout à la fois au vocabulaire, à la phraséologie, aux sigles des écrits communautaires pour en dénoncer le caractère cryptique et la pseudo-technicité. Ceux-ci véhiculeraient des notions aux contours volontairement mal circonscrits, dissimuleraient la vérité au citoyen et n'inciteraient pas l'opinion publique à s'intéresser à l'Europe.

Il n'est pas sûr que les censeurs ne visent pas plutôt la construction européenne que les mots créés pour la dire. À tout le moins font-ils porter aux mots tous les péchés de l'Europe. Pour leur part, les économistes et les juristes, spécialistes austères du langage plutôt glacé des chancelleries nationales, admettent, il est vrai du bout des lèvres, que les notions mises en jeu pour la construction européenne devaient bien se tailler un nouvel habit linguistique.

Critiquer ce langage n'est pas un crime de lèse-europe ; aucun langage n'est, par nature ou *de jure*, au-dessus de tout soupçon.

Cet état des choses oblige le linguiste à se poser plusieurs questions fondamentales. Quelle est la nature de ce langage et comment s'est-il instauré ? Quelle est son origine et quels sont ses modes de formation ?

Nous avons pour notre part engagé (Goffin 1994 : 636-642), voici quelques années, une réflexion générale sur la nature de ce (techno) lecte spécifique que s'est forgé et se forge encore l'UE pour décrire l'ordre juridique autonome auquel elle a donné naissance. Ce lecte, nous l'avons baptisé *eurolecte* (Goffin 1990 : 18). Il vise à dénommer les concepts en devenir et à maîtriser l'afflux d'unités lexicales nouvelles (*eurolexies*) d'une Europe en pleine construction.

La terminologie strictement eurolectale n'a été jusqu'ici ni recensée, ni dénombrée. Elle a certes été colligée, mais par excès. Il serait, en effet, pour le moins hasardeux de conférer le statut d'*eurolexie* à tel ou tel terme pour la seule raison qu'il figure dans un écrit communautaire ou qu'il s'y trouve actualisé. En accordant un peu légèrement l'investiture européenne à un ensemble flou de termes de nature économique, juridique ou administrative, la plupart des publications, telles que la série des *Glossaires des Communautés européennes*, éditée par le Conseil (1982), le *Wörterverzeichnis* (1984), publié par le Comité économique et social et le *Vocabulaire de base des C.E.* (1982) publié par la Commission et le *Europa-Wörterbuch*, préparé par K.-H. Kellner, Bruxelles, 1998, désorientent plutôt celui qui cherche à recenser ce vocabulaire. Quant aux nombreux alphabétiques consacrés aux institutions européennes, du *Dictionnaire de l'Européen* au *Guide to the Maze* en passant par *Parlez-vous eurocrate ?* – et qui envahissent le marché du livre partout et dans toutes les langues, leurs auteurs y rassemblent un cortège bien hétéroclite de termes du « langage communautaire » où se bousculent pêle-mêle des termes spécifiques comme *subsidiarité*, *prix de seuil* ou *stabilisateur agricole*, des sigles, des noms de programmes et de comités ainsi que des données biographiques et historiques sur les grands Européens (voir bibliographie).

2. RECENSEMENTS TERMINOLOGIQUES. MATÉRIAUX TEXTUELS OU *CORPUS* SUR PAPIER

La construction européenne a donné lieu à une législation abondante, sinon débridée. Les écrits communautaires constituent un vaste ensemble de textes instituant le droit originaire (primaire) et le droit dérivé (droit secondaire) et des textes décrivant les politiques communes. Comme la Communauté (UE) est une création de droit, la majorité

des écrits qu'elle produit relèvent du mode performatif, c'est-à-dire que les textes ne décrivent pas la réalité mais la créent.

Parmi les textes du droit originaire, on distingue : le traité de Paris (1951), le traité de Rome (1957), l'Acte unique (1987), le traité sur l'Union européenne, dit *de Maastricht* (1991), le traité d'Amsterdam (1992), le traité de Nice (2001) et le projet de Constitution européenne (2005).

Vu le caractère normalisateur du discours législatif communautaire, ces textes ont acquis la dignité des textes classiques et les termes qu'ils véhiculent le statut de termes canoniques.

Ils font partie de l'*acquis communautaire* (80 000 pp.) qui correspond au socle commun de droits et d'obligations qui lie l'ensemble des États membres au titre de l'Union européenne. Ils englobent les textes de mise en œuvre d'une politique commune dans le domaine agricole, économique, de la concurrence et des transports, celles de l'environnement, de la politique sociale, régionale et de l'union monétaire (introduction de l'Euro).

La première organisation supranationale européenne CECA vit le jour en 1951, ayant pour mission de fonder les « premières assises d'une communauté plus large et plus profonde » grâce à la mise en place d'un marché commun du charbon et de l'acier. Dans le traité de Paris, on trouve par exemple des termes tels que *subventions et aides, consultation, Haute Autorité, notification, sanctions pécuniaires et astreintes, prélèvement, mesures conservatoires*.

En 1957, les États membres signèrent à Rome les traités instituant la CEE et l'Euratom. Dans le traité de Rome et dans le droit dérivé, on trouve entre autres *prélèvements (agricoles), systèmes de préférences généralisées, prix de seuil, mesures d'effet équivalent, montants compensatoires monétaires, prélèvement de coresponsabilité, ressources propres*.

En 1987, l'Acte unique européen modifiait et complétait les traités de Rome, prévoyant pour 1992 l'achèvement du marché intérieur et introduisant la *procédure de coopération*. Le terme « unique » qualifiait l'Acte parce qu'il s'agissait de réunir dans un même cadre juridique les communautés européennes, la CPE et l'AUE.

Le Traité d'Amsterdam de 1997 et celui de Nice de 2001 ont été animés par une volonté politique plus faible. C'est à Amsterdam que fut introduit le concept de *coopération renforcée*. Une révision plus complète des traités fut signée dans le Traité sur l'Union européenne de Maastricht (1992). Il contient des innovations institutionnelles, *procédure de codécision, principe de subsidiarité*, instaure une *citoyenneté de l'Union* et prévoit des *régimes dérogatoires fondamentaux*.

L'encre de la controverse n'a pas fini de couler à propos du projet de Constitution européenne. Il y eut d'entrée de jeu une polémique sur le titre ; s'agit-il d'un traité (*Vertrag, Treaty*), d'une Constitution (*Verfassung, Constitution*) ou d'un Traité constitutionnel (*Verfassungsvertrag*) ? Apparaissent pour la première fois dans la Constitution les lexies : *principe de proportionnalité, mécanisme de l'alerte précoce, disposition de*

passerelle, bonne gouvernance, concurrence libre et non faussée (disposition qui date de 1957), *service d'intérêt général*.

3. RECENSEMENTS TERMINOLOGIQUES. *CORPUS* INFORMATISÉS

Nous retenons cinq *corpus* complémentaires, à savoir :

1. SCAD
2. EURLEX
3. EURODICAUTOM
4. MULTITERM
5. EURAMIS

SCAD plus (*Service central automatisé de documentation*)

Ce site propose une synthèse de la législation de l'UE sous la forme de fiches d'information conviviales, réparties en 32 domaines thématiques correspondant aux activités de l'Union européenne. On y trouve non seulement des synthèses de la législation existante, mais également un suivi des propositions législatives dans des domaines aussi divers que les relations extérieures ou l'emploi et les affaires sociales. Avec près de 2 500 fiches d'information mises à jour quotidiennement, tous les domaines sont abordés, collant au plus près de l'actualité législative. Un millier de fiches contenant des informations d'intérêt général sont disponibles dans les onze langues qui étaient les langues officielles de l'Union européenne avant le 1^{er} mai 2004. Les informations plus spécialisées sont disponibles uniquement dans quatre langues de l'UE : l'anglais, le français, l'allemand et l'espagnol. Des outils tels que l'index par mots clés et le glossaire sont particulièrement précieux.

SCAD plus (<http://europa.eu.int/scadplus>) est riche de 166 000 pages d'information. Une page internet correspond à une seule adresse et peut comporter une page DIN A4 ou 100.

Le Glossaire de SCAD plus est accessible via http://europa.eu.int/scadplus/glossary/index_fr.htm.

Il présente 220 termes relatifs à la construction européenne, aux institutions et activités de l'Union européenne.

La base de données juridiques EURLEX (*ancien Celex*)

EURLEX (*Communitatis europae lex*) est la base de données juridiques officielle des institutions de l'Union européenne. Elle recense la législation communautaire, y compris les travaux préparatoires et les références aux mesures nationales d'exécution, la jurisprudence de la Cour de justice, les questions parlementaires. Elle est disponible dans neuf langues.

Sa conception documentaire la rend très différente d'Eurodicautom, d'autant qu'elle joint le texte intégral à une structure analytique très riche. Eurlex est riche de 1 060 000 pages.

EURODICAUTOM

Eurodicautom est la banque de données terminologique riche de quatre millions d'entrées en plusieurs langues (Goffin 1999 : 185-224).

MULTITERM

Multiterm '95 Plus, un progiciel applicatif de gestion de données terminologiques multilingues, a été installé sur les PC des traducteurs début 1996 et vise la valorisation au profit de tous les traducteurs de la terminologie rassemblée dans les diverses Unités de traduction. Un système de filtres (attributs) combinables avec des recherches par champs indexés ou par chaîne de caractères, permet de trier les données.

Les mémoires de traduction : le projet EURAMIS

Le projet *EURAMIS* (European Multilingual Information System) s'appuie sur la constatation que le Service de traduction dispose dans ses archives d'un vaste *corpus* de données textuelles parallèles dans toutes les langues officielles et le fait que le nombre de textes répétitifs est non négligeable : d'où la tentation de constituer une « mémoire de traductions ». Cette mémoire permet de tirer profit de la répétitivité par la réutilisation de textes ou parties de textes déjà traduits, de correspondances terminologiques, voire de références aux documents antérieurs. Toutefois, ces ressources linguistiques ne peuvent être exploitées utilement que si l'outil procède à l'alignement du texte source et du texte cible pour faire apparaître les correspondances entre langues, qu'il stocke les phrases ainsi alignées, qu'il permette la consultation phrase par phrase, par terme ou morceau de phrase.

4. TYPOLOGIE

Les mécanismes lexicogéniques qui sous-tendent l'eurolecte sont, pour une large part, comparables à ceux que la linguistique a mis en évidence pour toute production néologique.

Comme nous l'avons montré ailleurs (Goffin 2002 : 167-178), nous distinguons trois types de néologismes :

- les eurolexies néologismes morphématiques, obtenues par dérivation affixale (6 %) : *comitologie, surbudgétisation, périphéricité, additionnalité, ultrapériphérique* ou obtenues par le morphème préfixal *euro*, qui se montre d'une belle vitalité, dans *euro-pêche, europium, eurobaromètre*, ou en allemand *Eurokaner, Eurokratur, Eurokrat* ;
- les eurolexies néologismes sémantiques (10 %) : dans ce cas une signification nouvelle est donnée à un segment phonologique existant. Le terme prend une significa-

sion supplémentaire par glissement ou transfert, la restriction à la sphère communautaire étant compensée par un enrichissement du contenu sémantique, par exemple : *règlement* (et son champ notionnel), *directive*, *recommandation*, *subsidiarité*, *collégialité*, *prélèvement*, *approfondissement*, *transparence*, *cabotage* ;

- les eurolexies néologismes syntagmatiques (51 %) résultant de la collocation, parfois en cascade, de déterminants successifs, présentant une unité de signification permanente.

On distingue ici :

- le syntagme asyndétique : *action monofonds*, *approche plurifonds* ;
- la collocation synthématique : substantif suivi de l'adjectif équivalent d'arguments : *acquis communautaire*, *coopération renforcée*, *livre blanc*, *livre vert* ;
- le syntagme à joncteur « à » ou « de » : Europe à géométrie variable, transport aérien de quatrième liberté, démonstration communautaire.
- On peut aussi les classer suivant leur origine ou provenance et on obtient :
 - les eurolexies « autochtones » ou endolexies, purement eurolectales, tels que *comitologie*, *acquis communautaire*, *abstention constructive*, *livre blanc*, *livre vert* ;
 - les eurolexies « allochtones » ou endolexies (hétérolexies) obtenues par transfert d'un terme appartenant à un autre vocabulaire de spécialité, p. ex. *médiateur*, *subsidiarité*, soit par transfert d'un terme appartenant à un vocabulaire général, p. ex. *troïka*, *élargissement*, *gouvernance*, *noyau dur*.

5. LA NATURE DU TEXTE COMMUNAUTAIRE (CONSIDÉRATIONS TEXTOLOGIQUES)

Par le jeu original de ses institutions, l'Union européenne combine, en une heureuse mixité, l'autorité intergouvernementale du Conseil de l'Union, qui représente les États-Nations, et l'autorité communautaire représentée par la Commission. Dans cette entité hermaphrodite, le processus décisionnel diffère sensiblement de celui qui dans les États emprunte la voie parlementaire puisque le processus communautaire vise à établir une norme commune à plusieurs États, susceptible de concilier des traditions différentes.

Il apparaît assez rapidement qu'un texte communautaire prend souvent appui sur un modèle national prépondérant, lequel marque de son empreinte le texte final, fruit d'un compromis, quant au fond (habitudes de pensée) et quant à la forme (habitudes de style ou des dictions « nationales »). Le texte du statut de la fonction publique européenne est ainsi, pour une large part, décalqué de la version française. La diction s'en ressent inévitablement. Ces textes font ainsi entrer en équilibre instable, les termes « nationaux » façonnés par une longue tradition juridique, sociale et culturelle, et les eurolexies ou euronymes qu'implique la mise en place d'une autorité communautaire supranationale. Il serait hasardeux d'affirmer qu'une coupure, factice et fâcheuse, oppo-

serait les notions nationales transparentes dans leur évidence et les notions supranationales opaques et pour d'aucuns volontairement cryptées.

6. LA DIMENSION CONTRASTIVE

6.1. Emprunts et calques

La terminologie différentielle a mis en perspective que chaque langue privilégie les schèmes de composition, et une combinatoire syntagmatique spécifique. Alors que le français construit des syntagmes épithétiques avec adjectifs de relation, comme *abstention constructive*, *intégration différenciée* ou des synapsies à ligament comme *critères de convergence*, *droit d'initiation*, *Europe à géométrie variable*, l'anglais use de diverses formes de concaténation, des séquences de segments disjoints avec trait d'union, comme *general-interest-services* ou des segments autonomes, tels que *codecision procedure*, *Union withdrawal clause* ; quant à l'allemand, il privilégie la composition telle que *Mitverantwortungsabgabe* (Fr. prélèvement de coresponsabilité).

Chaque langue paraît donc suffisamment riche pour faire face, à son rythme verbal et suivant ses modèles, aux défis néologiques d'aujourd'hui. Dans chaque langue tout peut être dit, mais tout n'a pas encore de nom.

Mentionnons ici deux aspects souvent privilégiés de la néologie de forme : l'emprunt et le calque. Ce sont presque toujours le français et l'anglais qui servent de langue source. Dans la terminologie liée au basculement vers l'euro, on trouve les emprunts à l'anglais, p. ex. *opting-out*, (option de sortie, clause d'exemption : dérogation accordée à un pays ne souhaitant pas se rallier aux autres États membres dans un domaine particulier).

Pre-in, *pays pre-in*, (pays qui ne participe pas d'emblée à l'union monétaire) ; *pre-out* : (État qui d'emblée ne fait pas partie de l'union monétaire).

Étant lui-même un procédé de transposition littérale (à preuve l'anglais *loan translation* et l'allemand *Lehnübersetzung*), le calque trouve dans la traduction un milieu d'extension favorable, puisqu'il s'agit d'une traduction en morphèmes autochtones, morphème par morphème d'une construction allogène. Ici, c'est le français qui tient le rôle de grande source et, à ce titre, imprègne et influence les autres langues au niveau lexical et sémantique. Tout en constatant l'emprise du français par le calque sur les autres langues romanes, certains linguistes espagnols l'accueillent favorablement au nom de la transparence et de l'internationalisation. D'autres, plus respectueux de la tradition linguistique que sensibles aux internationalismes, craignent que le modèle étranger ne vienne déstabiliser les structures spécifiques de la langue réceptrice.

6.2. Texte original et traductions

À l'image du texte-compromis, les versions traduites laissent transparaître en filigrane les modes d'expression du modèle, tant est forte la prégnance du texte original. Sans être asservie à ce modèle, la traduction s'en empreint de manière diffuse mais réelle, obligée de se plier, par souci d'exactitude, à ses contours et à son architecture. Cette

instillation se trouve renforcée par la publication concomitante de textes, parfois segment par segment. De valeur juridique égale, les versions sont donc, du point de vue strictement linguistique, moins authentiques et moins autochtones que le modèle original. Comme, à la Commission, le français tient souvent le rôle de langue originale, les traductions se glissent dans le moule du texte préconçu quand bien même sa diction originelle serait tempérée par un travail « transnational » des juristes-linguistes.

Compte tenu de ce modèle sous-jacent, il faut du temps pour apprendre à déchiffrer un texte communautaire et à pouvoir reconnaître, en filigrane, la facture originale et les indices dont la version finale est porteuse. L'examen des moutures successives d'un même texte permettrait d'en reconstituer la genèse « philologique » en opposant au texte fini un ou des avant-textes (génétique textuelle).

7. LISIBILITÉ ET CONSOLIDATION

Le manque de transparence a souvent été évoqué dans le cadre communautaire, où il implique un appel à la clarté de fonctionnement des institutions et à une meilleure lisibilité des textes. L'Europe est même ressentie de façon généralisée comme lointaine et secrète, cachée par des procédures de décision difficiles à comprendre.

Or, assurer aux Européens une législation communautaire plus concise et compréhensible, rédigée d'une « plume plus légère » (J. Delors), dans une langue d'accès plus aisé, est un souhait explicitement formulé dans les Conclusions de la Présidence, au sommet d'Edimbourg, le 12 décembre 1992. Celles-ci n'hésitent pas à décrire les deux démarches possibles : la consolidation et la codification.

L'édifice juridique communautaire est le fruit de modifications successives des traités et autres documents officiels, lesquelles se reflètent dans la structure alambiquée et une organisation parfois touffue des articles. Comme on peut le lire dans le glossaire « Réforme de l'Union européenne » (voir *bibliographie*), « la consolidation des textes législatifs correspond à une simplification purement déclaratoire et officieuse d'actes juridiques ». C'est l'intégration des diverses modifications dans l'acte de base, une opération de clarification menée par la Commission.

La publication *traité sur l'Union européenne* reproduit ainsi le texte du traité tel qu'il a été signé le 7 février 1992 à Maastricht. Il est complexe dans sa structure, avec des articles A, B, C (en lettres), puis des insertions article 3A, des remplacements et des modifications, d'où un effet d'opacité. Le texte publié en 2003 contient la version dite *consolidée* de ce traité dans laquelle se trouvent intégrées les modifications apportées par le traité de Nice.

La comparaison des deux textes laisse apparaître, dans la version consolidée, un souci manifeste de clarté et de précision.

8. L'EUROLECTE DANS LA DYNAMIQUE DU DISCOURS

Si l'on est prêt à reconnaître la spécificité de la construction européenne appelée à créer un ordre juridique nouveau dans le respect des traditions nationales, il faut admettre du

même coup que le langage qu'elle se forge et qu'elle manie doit être replacé dans le processus d'intégration en cours.

Qu'il donne lieu à la création de mots nouveaux (*euronymes*, eurolexies néonymiques) pour désigner des concepts nouveaux ou en devenir, qu'il transmette des notions et des désignations enracinées dans l'une ou l'autre des langues protagonistes (*hétérolexies*), ce langage est constitué d'un vaste ensemble d'énoncés qui fonctionnent comme un sous-système au sein d'une langue et non pas comme un langage artificiel, ni un jargon, forme dégradée ou corrompue, d'une langue « nationale », elle-même claire et dépouillée.

De par sa nature, ses origines, ses modes de formation et son fonctionnement, ce technolecte – auquel on peut conférer la dignité d'*eurolecte* – ne se démarque aucunement des mécanismes qui gouvernent toute langue de spécialité. Une fois replacées dans la logique des énoncés, réinsérées dans leur contexte cognitif et réinscrites dans la dynamique du discours dans laquelle elles fonctionnent, les eurolexies prennent des allures moins ésotériques et se révèlent plus motivées et plus transparentes, si l'on prend en considération qu'elles émanent d'une pratique sociale de communication, qui au-delà de la convergence des économies et des politiques, respecte la variété des dimensions humaines et des valeurs culturelles.

BIBLIOGRAPHIE

- Budd (S.), 1987 : *The EEC. A guide to the Maze*, Londres, Kogan.
- Ceneco, 1991 : *Dixeco de l'Europe*, Paris, Dunod.
- Cloos (J.), Reinesce (G.), Vignes (D.) et Weyland (J.), 1993 : *Le Traité de Maastricht, Genèse, Analyse et commentaires*, Bruxelles, Bruylant.
- Crampton (S.), 1992 : *Eurospeak explained*, Londres.
- Doutriaux (Y.), 1992 : *Le Traité de l'Union européenne*, Paris, Colin.
- EC-Wörterverzeichnis*, 1984, français-allemand, Bruxelles, CES.
- European Communities Glossary*, 1990, français-anglais. Publication Conseil CE.
- Glossaire : La réforme de l'Union européenne en 150 définitions*, 1997, Opoce, Luxembourg (toutes langues) ; 2^e éd. 2005, 200 termes.
- Goffin (R.), 1990 : « L'Europe en neuf langues : champ d'affrontements et ferment d'intégration linguistique », dans *Meta*, 35-1, pp. 13-19.
- Goffin (R.), 1994 : « L'Eurolecte : oui, jargon communautaire : non », dans *Meta*, Hommage à B. Quemada, 39-4, pp. 636-642.
- Goffin (R.), 1998 : « Eurolekt ja, Eurokauderwelsch nein », dans *Deutscher Terminologie – Tag e.v. Akten des Symposiums Deutsche Terminologie im internationalen Wettbewerb*, Cologne, pp. 165-176.
- Goffin (R.), 1999 : « La banque EURODICAUTOM. La banque de données terminologiques multilingues de la Commission européenne 1973-1997 », dans *Cahiers de lexicologie*, n° 74, 1999-1, pp. 185-224.
- Goffin (R.), 2000 : « Le poète europarlamentaire et la galerie à traductrices ». À propos du poème de Marcel Thiry « Boulevard de l'empereur », dans *Il fabbro dela parlar materna*. Hommage à J.-M. Van der Meerschen, Éd. du Hazard, Bruxelles, pp. 103-113.

- Goffin (R.), 2002 : « Eurolecte : Analyse contrastive de quinze eurolexies néologiques » dans *Cahiers de lexicologie*, 2002-1, n° 80, pp. 167-177.
- Gontrand (F.), 1991 : *Parlez-vous eurocrate ?* Paris, Éd. Organisation, traduit par P. Bowen, *Eurospeak, A User's Guide*, Londres, Brealy.
- Kellner (K.-H.), 1998, *Europa-Wörterbuch, Französisch-Deutsch*, Teubner (R.), Verlag Brüssel Ottenbach.
- Paxton (J.), 1986 : *Dicionario das Comunidades Europeias*, Bruxelles.
- Projet de Traité établissant une Constitution pour l'Europe* : juillet 2003, version consolidée, juillet 2004.
- Svaldi (A.), 1997 : *Terminologia comunitaria*, Roma, Istituto di tecnica e propaganda agraria.
- Thøgersen, (N.), 1988 : *700 Ord om EF*, publication communautaire.
- Visine, (F.), 1987 : *Dictionnaire de l'Européen*, Paris, Visine.
- Vocabulaire de base de la C.E.* (1992), anglais-français-chinois, Publication CE, Commission.
- Weidenfeld (W.), Wessels (W.), 1991 : *Europa von A-Z*, Bonn, Europa Union Verlag.
- Zerwes (H.), 1992 : *Referenz-Glossar für Legaldefinitionen* (4 langues), Comité économique et social.

LES SPÉCIFICITÉS EN CONTEXTE : COMMENT ÉTUDIER LA POLYSÉMIE DANS UN *CORPUS* TECHNIQUE ?

Ann Bertels

Institut interfacultaire des langues vivantes (ILT)

Quantitative Lexicology and Variational Linguistics (QLVL)

K.U. Leuven

Louvain, Belgique

ann.bertels@ilt.kuleuven.be

Résumé : Nous présentons dans cet article les principes méthodologiques et les premiers résultats d'une étude sémantique quantitative du vocabulaire spécifique d'un corpus en français technique. Les spécificités seront identifiées avec la méthode des mots-clés (Keywords Method), en comparant le corpus technique à un corpus de référence. Le contexte des spécificités, et plus particulièrement le recouvrement des cooccurrences de leurs cooccurrences, permettra ensuite de vérifier si et dans quelle mesure les spécificités dans ce domaine technique sont monosémiques ou polysémiques.

Mots-clés : langue de spécialité, spécificités, polysémie, cooccurrences

1. JUSTIFICATION MÉTHODOLOGIQUE

Cet article s'inscrit dans le cadre d'une thèse de doctorat qui est une étude sémantique quantitative du vocabulaire spécifique d'un *corpus* en français technique. Les textes du *corpus* relèvent du domaine technique des machines-outils pour l'usinage des métaux. Grâce aux *corpus* électroniques de textes spécialisés, on dispose d'une information particulièrement intéressante pour l'analyse sémantique, à savoir le contexte linguistique. Toutefois, l'exploitation de grandes quantités de textes requiert une approche quantitative et automatisée.

Pour étudier la sémantique dans la langue de spécialité, l'approche qu'on adoptait traditionnellement est une approche catégorielle. Cette approche catégorielle se caractérise, d'une part, par un classement catégoriel ou binaire des unités linguistiques, c'est-à-dire par la dichotomie termes-mots. D'autre part, elle se caractérise par un classement catégoriel ou binaire au niveau sémantique, à savoir l'opposition monosémie – polysémie. Alors que la terminologie traditionnelle préconise la monosémie et l'univocité des termes d'une langue de spécialité, nous proposons d'adopter, pour notre étude sémantique quantitative et automatisée, une approche scalaire. Au classement catégoriel des unités linguistiques (termes-mots), nous opposons un *continuum* de

spécificité¹, allant des unités linguistiques les plus spécifiques aux moins spécifiques. Le classement catégoriel au niveau sémantique (monosémie vs polysémie) est remplacé par un continuum sémantique de monosémie, allant des unités plus monosémiques aux moins monosémiques ou, ce qui revient au même, plus polysémiques. Afin de vérifier ou de falsifier la thèse des monosémistes traditionnels, nous procédons donc à des évaluations fondées sur cette double dimension et à des analyses statistiques, ce qui revient à une double analyse quantitative (celle des spécificités et celle de la monosémie).

Pourquoi cette approche par *continuum* ? La linguistique de *corpus* ainsi que l'analyse de *corpus* spécialisés ont remis en question l'approche onomasiologique prescriptive et normative de la terminologie traditionnelle et ont contribué à l'essor de la terminologie descriptive. Sur le plan des unités linguistiques, les partisans de la terminologie descriptive rejettent la dichotomie entre langue générale et langue spécialisée et adoptent une approche sémasiologique et linguistique (Condamines et Rebeyrolle 1997). Les termes font partie intégrante de la langue naturelle, mais ils se caractérisent par le fait qu'ils véhiculent des connaissances spécialisées (Lerat 1995). En plus, le vocabulaire d'un *corpus* technique ne contient pas uniquement des mots techniques ou « termes » au sens strict, qui sont propres au domaine de spécialité, tels que *usinage* ou *broche*, mais également des mots du VGOS (vocabulaire général d'orientation scientifique) (Phal 1971). Ces mots s'emploient dans plusieurs domaines scientifiques et techniques et leur sens est déterminé par les contextes spécialisés (*machine, outil*). Finalement, le vocabulaire d'un *corpus* technique comprend également des mots de la langue générale, tels que *type, modèle, permettre*, etc. Sur le plan de l'analyse sémantique, la monosémie et l'univocité de la langue de spécialité ont été remises en question notamment par la théorie communicative de la terminologie (Cabré 1998 et 2000), par la socioterminologie (Gaudin 1993) et par la terminologie sociocognitive (Temmerman 1997 et 2000). Ces remises en questions nous incitent à adopter une approche par *continuum*, aussi bien au niveau des unités linguistiques (*continuum* de spécificité) qu'au niveau sémantique (*continuum* de sens ou *continuum* de monosémie).

2. OBJECTIFS ET PROCÉDURE

Pour l'analyse sémantique du *corpus* technique, nous procédons à une double analyse quantitative et nous articulons notre question de recherche autour de deux axes méthodologiques. Dans un premier temps, nous déterminons le vocabulaire spécifique du *corpus* technique (= les spécificités) et dans un deuxième temps, nous calculons à partir du contexte le degré de monosémie ou polysémie de ces spécificités, en implémentant la monosémie comme l'homogénéité sémantique. Comme nous proposons de quantifier la recherche, nous nous fixons pour but de vérifier s'il y a une corrélation entre, d'une part, le *continuum* de spécificité et, d'autre part, le *continuum* de monosémie

¹ Nous adoptons le terme « spécificités » pour désigner les mots les plus spécifiques et caractéristiques du corpus technique d'analyse, indépendamment de la méthode utilisée (calcul des spécificités vs *Keywords Method*).

(*continuum* de sens). Cette double approche par continuum requiert bien entendu une double expression de degré : le degré de spécificité et le degré de monosémie. Les mots du *corpus* technique seront donc ordonnés en fonction de leur degré de spécificité et situés sur une échelle de spécificité (*continuum* de spécificité). Le degré de monosémie permettra ensuite de situer les mêmes mots spécifiques sur une échelle de monosémie (*continuum* de monosémie ou *continuum* de sens), à partir d'une analyse des cooccurrences de deuxième ordre.

Notre double approche quantitative par *continuum* nous amène à poser une question scientifique de manière mesurable et objective. La question principale est celle de savoir s'il y a une corrélation entre le rang de spécificité et le rang de monosémie. Nous avançons l'hypothèse que les mots (les plus) spécifiques du *corpus* technique ne sont pas nécessairement (les plus) monosémiques. Citons par exemple le mot *tour* : (1) « machine-outil pour l'usinage des pièces », (2) « rotation, révolution » et (3) un sens général dans *attendre son tour/à son tour*. La question principale sera complétée par d'autres questions faisant intervenir les différents facteurs qui influent sur le degré de monosémie, notamment la fréquence technique, la fréquence générale, la classe lexicale, la longueur et le nombre de classes lexicales. Une analyse de régression multiple permettra de vérifier l'impact des variables indépendantes sur le rang de monosémie (variable dépendante).

3. CORPUS TECHNIQUE D'ANALYSE

Le *corpus* technique d'analyse est constitué de textes techniques spécialisés, relevant du domaine technique des machines-outils pour l'usinage des métaux. Il comprend environ 1 760 000 mots et il a été étiqueté et lemmatisé par *Cordial 7 Analyseur*. Le *corpus* technique consiste en quatre sous-*corpus*, datant de 1998 à 2001 : revues techniques électroniques (800 000), fiches techniques (300 000) trouvées sur Internet, normes Iso et directives (300 000) et quatre manuels numérisés (360 000). Les quatre sous-*corpus* se situent à différents niveaux de normalisation et de vulgarisation, assurant ainsi la représentativité du *corpus*. Pour pouvoir identifier les spécificités, le *corpus* technique d'analyse est complété par un *corpus* de référence de langue générale, constitué d'articles journalistiques du journal *Le Monde* (janvier-septembre 1998). Le *corpus* de référence a également été étiqueté et lemmatisé et comprend environ 15 300 000 mots (rapport de 1 à 10).

4. IDENTIFIER LES SPÉCIFICITÉS

L'analyse sémantique quantitative s'articule autour de deux axes méthodologiques : premièrement l'identification des spécificités et deuxièmement le calcul du degré de monosémie ou polysémie. Le premier axe méthodologique est l'axe des spécificités. Au classement catégoriel (termes – mots), nous opposons un *continuum* de spécificité, où les mots-clés sont situés en fonction de leur degré de spécificité. Les spécificités ou mots-clés (*keywords*) ne sont pas les mots les plus fréquents du *corpus* technique d'analyse : ce sont en revanche les mots les plus représentatifs, qui caractérisent thématiquement le *corpus* technique d'analyse. En termes relatifs, les spécificités sont signifi-

ficativement plus fréquentes dans le *corpus* spécialisé que dans un *corpus* de référence de langue générale. Afin de déterminer les spécificités, plusieurs méthodes sont désormais disponibles. La *keywords method* ou méthode des mots-clés est basée sur la statistique du LLR (*log likelihood ratio*) (log de vraisemblance) et implémentée dans les logiciels *Abundantia Verborum Frequency List Tool*² et dans *WordSmith Tools*³. La deuxième méthode du calcul des spécificités est basée sur la distribution hypergéométrique et elle est implémentée dans le logiciel *Lexico3*⁴.

Nous utilisons la *keywords method*, qui compare les fréquences dans le *corpus* spécialisé (*corpus* technique d'analyse) aux fréquences dans le *corpus* de référence, compte tenu de la taille des deux *corpus*. Ceci revient à comparer la fréquence observée (dans le *corpus* technique d'analyse) à la fréquence attendue (dans le *corpus* de référence). Si la fréquence observée est supérieure à la fréquence attendue et si cette différence est statistiquement significative, les mots sont spécifiques du *corpus* technique et sont donc considérés comme des spécificités positives. Après suppression des *hapax* et après avoir filtré les mots grammaticaux et les noms propres, nous avons recensé environ 5 000 spécificités dans le *corpus* technique, tant des termes au sens strict tels que *fraisage*, *alésage* que des mots tels que *type*, *permettre*. La mesure statistique du LLR (log de vraisemblance) indique le degré de spécificité et permet de classer les mots-clés par ordre de spécificité décroissante et par conséquent, de les situer sur un *continuum* à partir de leur rang de spécificité.

² *Abundantia Verborum* : <http://www.ling.arts.kuleuven.be/genling/abundant>

³ *WordSmith Tools* version 3 : <http://www.lexically.net/wordsmith/> et <http://www.oup.com>

⁴ *Lexico3* : SYLED – CLA2T, Paris3 : <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexico>

4. Spécificités

TOKEN	ABSFREQ1	ABSFREQ2	LIKESTAT	1-P	RELFREQ1	RELFREQ2	COM
machine	12671	1052	50521,8994	1	74,5114066	0,70934381	1
outil	8306	918	32037,7132	1	48,843165	0,61899013	1
usinage	6720	8	30468,408	1	39,5167432	0,00539425	1
pièce	7556	2219	24407,4535	1	44,4328142	1,49622996	1
mm	5490	191	23357,569	1	32,2837679	0,12878771	1
vitesse	5283	900	19108,7716	1	31,066511	0,60685307	1
coupe	6730	4153	17063,3653	1	39,5755478	2,80028978	1
broche	2893	12	13010,4136	1	17,0121932	0,00809137	1
Fig	2680	0	12193,9983	1	15,7596535	0	1
axe	3206	420	12079,1616	1	18,8527796	0,2831981	1
copeau	2557	0	11634,1809	1	15,036356	0	1
plaquette	2407	35	10592,4549	1	14,1542858	0,02359984	1
diamètre	2415	95	10200,0859	1	14,2013296	0,06405671	1
commande	2751	850	8765,70542	1	16,1771667	0,57313901	1

Tête de liste (4717) : les mots les plus spécifiques du corpus technique



Figure 1 : Tête de liste des spécificités : les mots les plus spécifiques du corpus technique d'analyse.

5. QUANTIFIER LA POLYSÉMIE À PARTIR DU CONTEXTE

Le deuxième axe méthodologique de l'analyse sémantique quantitative est l'axe sémantique du *continuum* de monosémie. En préparation de l'analyse sémantique quantitative, nous avons mené une expérimentation sur un petit échantillon de 30 termes techniques, comprenant entre autres *broche*, *avance*, *découpe*. Les dictionnaires techniques spécialisés et l'étude du contexte linguistique à partir des concordances nous ont permis, lors de cette analyse sémantique manuelle, d'accéder au (x) sens de ces 30 termes et de constater leur polysémie au sein du *corpus* technique. Toutefois, pour l'analyse sémantique des 5 000 spécificités, l'automatisation et la quantification s'imposent. Nous avons implémenté la monosémie comme l'homogénéité sémantique et procédé ainsi à une première tentative d'opérationnalisation du degré de monosémie à partir des cooccurrences des cooccurrences. Comme les cooccurrents sont des désambiguïsateurs puissants selon Schütze (1998) et Véronis (2003), ils sont très précieux pour l'analyse sémantique automatisée et quantitative.

Mais pourquoi recourir aux cooccurrents des cooccurrents ou cooccurrents de deuxième ordre ? Le caractère monosémique ou polysémique d'une unité linguistique se caractérise par des contextes sémantiquement homogènes, si elle est monosémique,

ou sémantiquement hétérogènes si elle est polysémique. L'accès à la sémantique des cooccurrents se fait par leurs cooccurrents, c'est-à-dire par les cooccurrents des cooccurrents et plus particulièrement par le recouvrement des cooccurrents de deuxième ordre. Si les cooccurrents de deuxième ordre ont un degré de recouvrement élevé, ils sont fortement partagés par les cooccurrents, ce qui est une indication de l'homogénéité sémantique des cooccurrents. Le degré de ressemblance ou de similarité lexicale des cooccurrents d'un mot de base étant proportionnel au degré de monosémie de ce mot, un fort recouvrement des cooccurrents de deuxième ordre signale ainsi un degré plus important de monosémie.

Par contre, si les cooccurrents de deuxième ordre sont formellement (très) différents, ils se recouvrent (très) peu et ils sont (très) peu partagés. Par conséquent, les différents cooccurrents sont sémantiquement plus diversifiés, car une structure formelle de cooccurrence différente indique un sens différent. Si les cooccurrents sémantiquement diversifiés appartiennent à plusieurs champs sémantiques, la spécificité aura moins de chances d'être monosémique. Pour la spécificité *tour*, on retrouve par exemple les cooccurrents significatifs indiqués dans les exemples [1] et [2], indiquant clairement deux sens différents de *tour*.

[1] La première est un tour sur lequel on usine l'alésage central...

= « machine-outil pour l'usinage des pièces »

[2] ... broches pouvant monter jusqu'à quinze mille tours par minute,...

= « rotation, révolution »

Afin de déterminer le degré de monosémie, nous proposons une formule de recouvrement basée sur le recouvrement formel des cooccurrents des cooccurrents (cc), tenant compte (1) de la fréquence d'un cc dans la liste des cc (= nombre de cooccurrents (c) apparaissant avec ce (cc)), (2) du nombre total de c et (3) du nombre total de cc. Un cc aura d'autant plus d'importance pour le recouvrement total qu'il figure plus souvent dans la liste des cc, c'est-à-dire si sa fréquence dans la liste des cc est plus élevée ou s'il est plus souvent partagé par les cooccurrents ou c.

$$\sum_{cc} \frac{fq\ cc}{\# \text{ total } c \cdot \# \text{ total } cc}$$

Formule de recouvrement des cooccurrents des cooccurrents

Considérons en guise d'exemple un cc fortement partagé : il est partagé par exemple par 5 c des 7 c au total. Cela veut dire que 5 c des 7 c apparaissent avec ce cc en question, ce qui indique un recouvrement important. Dès lors, nous proposons d'inclure dans le numérateur le nombre de c qui ont ce cc en commun (fq cc), par exemple 5, et d'inclure dans le dénominateur le nombre total de c, par exemple 7. Le recouvrement est donc exprimé par la fraction 5/7. En exprimant pour chaque cc le recouvrement par

la fraction *nombre de c avec le cc* (ou *fq cc*) divisé par *nombre total de c*, le résultat se situe toujours entre 0 (pas ou peu de recouvrement) et 1 (recouvrement important) et par conséquent, le résultat sera facilement interprétable. Si l'on somme pour tous les cc, il faut ajouter dans le dénominateur le nombre total de cc (au niveau des *tokens*), car on considère tous les cc, avec les doublons responsables du recouvrement. Il ne s'agit pas du nombre de cc différents, mais du nombre total de cc, à savoir tous les mots qui cooccurrent avec tous les c.

Remarquons que nous considérons les c et les cc au niveau des formes graphiques et non pas au niveau des lemmes (formes canoniques). De cette façon, la mesure de recouvrement permet de faire la distinction entre, par exemple, *pièce usinée* et *pièce à usiner*.

La mesure d'association utilisée pour déterminer les cooccurrences pertinentes est la statistique du LLR (log de vraisemblance). Le seuil de significativité étant très sévère (valeur $p \leq 0,0001$), on est sûr de relever uniquement les cooccurrences sémantiquement pertinentes. L'algorithme et les scripts en *Python*⁵ permettent de définir les paramètres (fenêtre d'observation, seuil de significativité, etc.) au niveau des cooccurrences (premier ordre) et au niveau des cooccurrences des cooccurrences (deuxième ordre). Les scripts génèrent une grande base de données avec toutes les informations statistiques pertinentes (LLR, seuil de significativité, etc.). La base de données indexée sera alors interrogée pour analyser le recouvrement des cooccurrents de deuxième ordre pour chaque spécificité comme base. À cet effet, la fonction *Python* de l'algorithme prévoit les paramètres suivants : la base (spécificité à analyser), le seuil de significativité pour les cooccurrents de premier niveau (p. ex. 0,9999 pour une valeur $p \leq 0,0001$), le seuil pour les cooccurrents de deuxième niveau et la base de données à interroger. Il y a plus de recouvrement, si plus de cooccurrents (c) partagent le même cc, ce qui signifie un poids plus lourd pour ce cc (score près de 1). Un cc moins/pas partagé indique donc peu/pas de recouvrement (score près de 0). Pour toutes les spécificités du *corpus* technique, nous pourrions ainsi calculer le degré de recouvrement et par conséquent le degré de monosémie. Ce degré de monosémie permet ensuite de situer les spécificités sur un *continuum* de monosémie (*continuum* de sens), à partir de leur rang de monosémie.

6. ANALYSES STATISTIQUES : PREMIERS RÉSULTATS

La question principale vise à vérifier la corrélation entre le rang de spécificité et le rang de monosémie. À l'aide d'une analyse de régression simple dans le logiciel statistique R⁶, nous évaluons l'impact du rang de spécificité (VI : variable indépendante ou explicative) sur le rang de monosémie (VD : variable dépendante ou expliquée). Le coefficient de corrélation Pearson (-0,71) indique une corrélation négative entre le rang de

⁵ <http://www.python.org/>

⁶ <http://www.r-project.org/>

spécificité et le rang de monosémie : les mots les plus spécifiques sont les moins monosémiques. L'analyse de régression simple est hautement significative ($p < 2,2e-16$) et le pourcentage de variation expliquée R^2 est de 51,57 %. La variation du rang de spécificité permet donc d'expliquer 51,57 % de la variation du rang de monosémie.

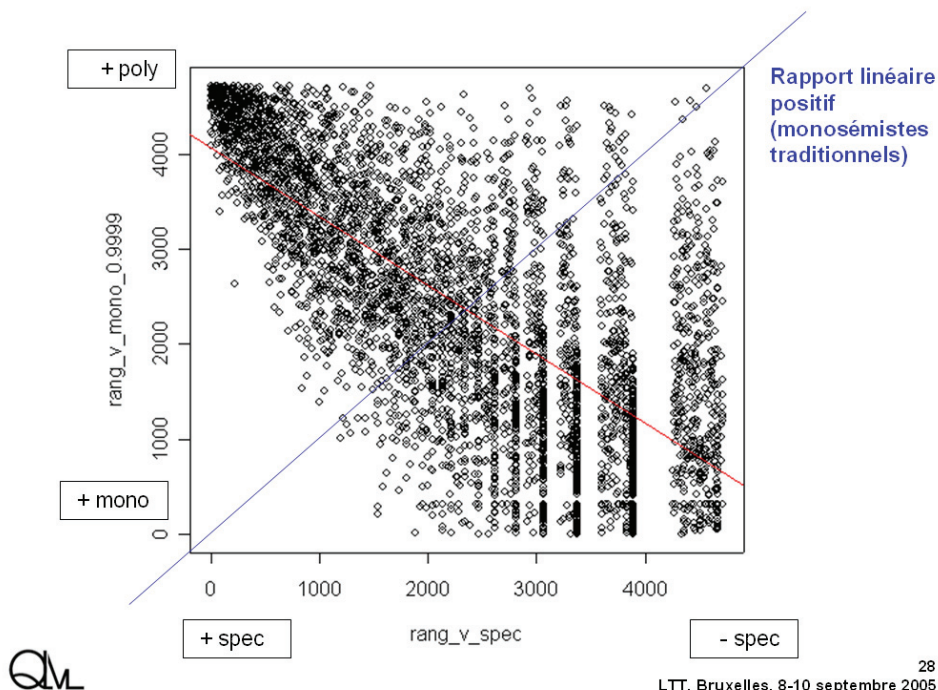


Figure 2 : Visualisation de l'analyse de régression simple : rang de spécificité (axe X) vs rang de monosémie au niveau 0.9999 (axe Y)

La visualisation ci-dessus (*cf.* figure 2) indique clairement la tendance linéaire négative, visualisée par la droite de régression descendante, contrairement au rapport linéaire positif des monosémistes traditionnels. Les résultats de l'analyse de régression simple ainsi que la visualisation permettent d'infirmer la thèse monosémiste évoquée plus haut. En effet, les mots les plus spécifiques de notre *corpus* technique d'analyse ne sont pas les plus monosémiques, mais au contraire, ce sont les plus polysémiques. En plus, les mots les moins spécifiques sont plutôt monosémiques, à quelques exceptions près, notamment *service*, *objet*, *commercial*, *eau*, *air*, etc. Ces exceptions à la tendance générale de corrélation négative sont des mots peu (ou très peu) spécifiques et polysémiques. Parmi les mots polysémiques, nous recensons donc les mots les plus spécifiques (*machine*, *outil*...), des mots moins spécifiques, ainsi que les mots les moins spécifiques (*service*, *objet*). Nous y retrouvons (1) des « termes » au sens strict ou des mots propres au domaine de spécialité, tels que *broche* et *usinage*, (2) le « VGOS » ou

vocabulaire général d'orientation scientifique, qui s'emploie dans plusieurs domaines scientifiques et techniques avec des mots tels que *machine* et *outil*. Les mots les plus polysémiques comprennent en plus (3) des mots que l'on pourrait qualifier de « mots de la langue générale » (*objet*, *eau*, *air*), ce qui est visualisé ci-dessous (cf. Figure 3).

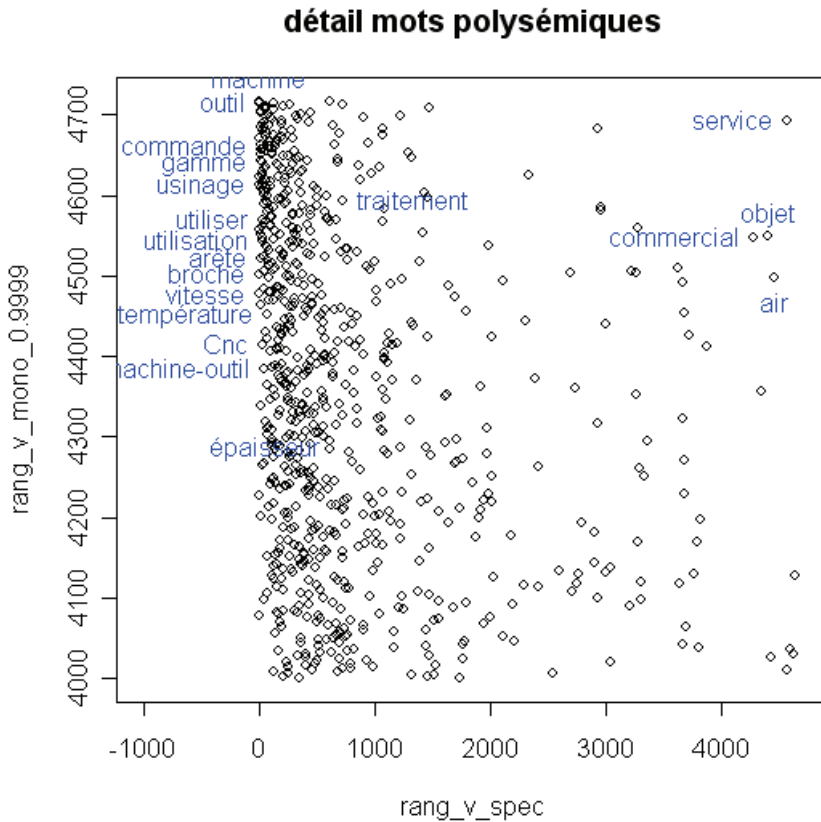


Figure 3 : Détail : visualisation des mots les plus polysémiques

7. RECHERCHES FUTURES

Des recherches supplémentaires s'imposent pour examiner la relation précise entre, d'une part, notre mesure de monosémie, implémentant la monosémie comme l'homogénéité sémantique, et, d'autre part, ce que l'on considère traditionnellement comme monosémie ou polysémie. Nous recourons à cette mesure de monosémie ou mesure de recouvrement, dans le but de développer un critère opérationnalisable et mesurable. Sans recherche supplémentaire, nous ne pourrions pas prétendre que notre mesure de monosémie correspond parfaitement à ce que les traditionalistes considèrent comme polysémie.

Comme le caractère monosémique ou polysémique d'une unité linguistique n'est pas uniquement influencé par sa spécificité, il est important d'analyser également l'influence d'autres facteurs influant sur le degré de monosémie, notamment la fréquence technique et générale, la classe lexicale, le nombre de classes lexicales, etc. Une analyse de régression multiple permettra d'évaluer l'impact combiné de toutes les variables indépendantes (VI) sur le rang de monosémie (VD).

Nous nous proposons également d'effectuer des analyses de régression simple et multiple détaillées, c'est-à-dire par classe lexicale (substantifs/adjectifs/verbes/adverbes) et par catégorie (par exemple pour les substantifs déverbaux, pour les abréviations et les sigles). En plus, il serait intéressant également de vérifier la corrélation entre le rang de spécificité et le rang de monosémie dans les différents sous-corpus (revues électroniques/fiches techniques/normes et directives/manuels). Ainsi, la question se pose de savoir s'il y aurait une meilleure corrélation ou peut-être une corrélation positive entre le rang de spécificité et le rang de monosémie dans les normes et directives, censées être prescriptives et normatives ?

Nous envisageons de procéder à une validation manuelle de la mesure de recouvrement à partir d'une analyse manuelle des collocations et nous travaillons actuellement à une mesure de recouvrement technique tenant compte de la spécificité des cooccurrents des cooccurrents dans le calcul du recouvrement. Ainsi, nous pourrions évaluer le degré de monosémie technique.

BIBLIOGRAPHIE

- Cabré (M.-T.), 1998 : *La Terminologie. Théorie, Méthode et Applications*, Ottawa/Paris, Presses de l'Université d'Ottawa/Armand Colin.
- Cabré (M.T.), 2000 : « Terminologie et linguistique : la théorie des portes » dans *Terminologies nouvelles*, n° 21, pp. 10-15.
- Condamines (A.), Rebeyrolle (J.), 1997 : « Point de vue en langue spécialisée », dans *Meta*, n° XLII-1, pp. 174-184.
- Gaudin (F.), 1993 : *Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles*, Rouen, Publications de l'Université de Rouen.
- Lerat (P.), 1995 : *Les Langues spécialisées*, Paris, PUF.
- Phal (A.), 1971 : *Vocabulaire général d'orientation scientifique (V.G.O.S.). Part du lexique commun dans l'expression scientifique*, Paris, CRÉDIF – Didier.
- Schütze (H.), 1998 : « Automatic Word Sense Discrimination » dans *Computational Linguistics*, n° 24-1, pp. 97-123.
- Temmerman (R.), 1997 : « Questioning the univocity ideal. The difference between socio-cognitive Terminology and traditional Terminology » dans *Hermes* n° 18, pp. 51-90.
- Temmerman (R.), 2000 : *Towards new ways of terminology description. The sociocognitive approach*, Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Véronis (J.), 2003 : « Cartographie lexicale pour la recherche d'informations » dans *Actes de TALN 2003*, pp. 265-274.

UN ASPECT DE L'APPORT DES *CORPUS* À LA TERMINOLOGIE LINGUISTIQUE : L'ALIGNEMENT

Franck Neveu

CRISCO (CNRS)

Université de Caen

Caen, France

Résumé : à travers les problèmes traductionnels liés à la non-coïncidence des deux métalangues grammaticales française et arabe, cette étude évoque les enseignements du recours aux contextes alignés pour la connaissance du fonctionnement du technolecte linguistique, et elle précise l'intérêt d'une approche textuelle de la terminologie.

Mots-clés : contexte, corpus, métalangue, terme complexe, terminologie linguistique

1. INTRODUCTION

L'objectif de cette étude est de préciser quelques-uns des principaux apports des *corpus* à la terminologie linguistique. L'intérêt de la question réside notamment dans le fait qu'une approche par les *corpus* (autrement dit par les textes et les discours) du technolecte linguistique est de nature à faire ressortir la spécificité terminologique non seulement du domaine linguistique mais aussi et plus largement de celui des sciences humaines.

L'exemplification de la problématique terminologique est fréquemment proposée à partir de domaines techniques et technologiques, voire empruntée aux sciences exactes et expérimentales. Travailler sur la question terminologique dans le domaine des sciences de l'esprit (dont relèvent, avec d'autres, les sciences du langage) fait nécessairement surgir des questions nouvelles et remet en cause quelques certitudes méthodologiques.

Toute science est un fait de culture, mais la particularité des sciences humaines est de prendre pour objets des faits de culture. Tandis que les objets des sciences exactes et expérimentales sont dans une relation d'extériorité avec l'observateur qui les étudie, cette extériorité n'est pas de même nature pour les sciences qui ont l'humain pour objet, qu'elles décrivent à la fois dans son unité et dans sa diversité historique et culturelle. De même, l'objectivité scientifique n'y est pas conceptualisée de la même manière. Le chercheur en sciences de l'homme participe toujours de son objet et ne saurait jamais s'en abstraire complètement quels que soient ses efforts ou ses prétentions méthodologiques. Cela n'est bien sûr pas sans conséquences sur l'activité terminologique, notamment pour ce qui concerne le développement du nombre des métatermes et leur variabilité définitionnelle et formelle. Les « objets » du langage ne sont pas des objets naturels.

Une partie du travail dont il va être ici question s'inscrit dans le cadre d'un programme scientifique soutenu par le réseau LTT de l'Agence universitaire de la Francophonie : *Étude contrastive de la métalangue grammaticale : terminologie et traduction*. Ce programme a pour objectif principal d'établir une terminologie linguistique bilingue français/arabe, à partir d'une analyse de *corpus*, en orientant plus particulièrement les recherches vers les notions de linguistique descriptive dans les domaines de la morphologie, de la syntaxe, et de la sémantique, sans exclure pour autant par principe les autres domaines¹. On bornera la présentation de ce projet aux questions soulevées par l'alignement.

2. SUR L'APPROCHE TEXTUELLE DE LA TERMINOLOGIE

Anne Condamines a récemment rappelé (2005) l'importance considérable prise aujourd'hui par l'approche textuelle de la terminologie, après la longue période wüstérienne d'une conception normative de la terminologie, normative par essence voire par objectif. La terminologie textuelle a connu un développement concomitant avec le développement de la linguistique de *corpus*, ce qui a rapproché la terminologie, par le biais du TAL, de la science du langage, dont la tradition normative la tenait jusque-là finalement assez éloignée. On construit aujourd'hui des terminologies à partir de *corpus*, et ce recours à l'étude de « données » terminologiques croise à la fois le besoin sans cesse réaffirmé d'empiricité en matière de description et d'explication linguistique, le besoin d'une exploration toujours plus avancée de la gestion électronique des documents et de l'ingénierie des connaissances, et bien sûr une forte demande sociale (notamment celle des entreprises qui ont à gérer, comme le rappelle A. Condamines, une documentation considérable, en lien avec la création, le développement et la maintenance d'objets manufacturés).

Monique Slodzian (2000), qui a étudié de manière approfondie l'approche textuelle de la terminologie, précise justement : « L'approche textuelle, infiniment mieux adaptée aux besoins et aux problèmes liés à une production effrénée de documents spécialisés, bouleverse les priorités. »

Alors que la terminologie classique restreint le linguistique aux seuls mécanismes de dénomination et impose une vue essentiellement paradigmatique, l'étude des textes de spécialité déplace l'intérêt sur :

¹ Dans le cadre de ce programme, les équipes françaises et tunisiennes se sont donné pour tâches principales : (i) de travailler à l'élaboration de ressources linguistiques, en mettant à la disposition des chercheurs et des enseignants qui ont à rédiger des articles ou des ouvrages de description grammaticale, des concordances et des textes alignés ; (ii) de proposer, sur la base des contextes rendus disponibles par ces corpus alignés, des définitions notionnelles susceptibles de pouvoir être recueillies dans le cadre d'un dictionnaire de terminologie linguistique bilingue français/arabe. L'objectif est de contribuer à la réflexion épistémologique sur la nature des métalangues grammaticales en confrontant deux systèmes linguistiques très différents : (i) celui d'une langue indo-européenne dont les descriptions grammaticales s'inscrivent dans la tradition occidentale, profondément marquée par la philosophie aristotélicienne, (ii) et une langue sémitique dont la tradition grammaticale reflète une pensée linguistique largement déterminée par l'interprétation du texte sacré qu'est le Coran.

- le fonctionnement réel des unités lexicales en contexte ;
- l'approche descriptive des textes et des unités lexicales au détriment de l'approche normative (la norme étant désormais considérée comme la résultante des contrôles successifs imposés sur les productions textuelles par le système axiologique du domaine et de la communauté) ;
- l'approche « bottom-up » pour la création d'ontologies à partir des textes ;
- la validation d'un réseau de termes issus d'un *corpus* donné par confrontation à un autre *corpus* analogue.

La terminologie textuelle, élaborée à partir de textes de spécialité réunis en *corpus*, permet en effet de construire des terminologies différenciées et des ontologies régionales. Elle conduit donc à repenser le rapport entre le syntagmatique et le paradigmatic, puisqu'elle procède inductivement à partir d'occurrences. Seule cette approche permet d'observer des phénomènes variationnels et de mesurer le décalage entre une forme lexicalisée et son usage.

M. Slodzian a montré notamment, à partir de travaux réalisés sur des *corpus* textuels représentatifs, que les unités produisent de la variation par insertion, par coordination, ou par permutation, et que ces faits de variation terminologique peuvent atteindre de 15 % à 25 % de la totalité des unités attestées comme termes.

Mais l'approche textuelle de la terminologie, par le biais de la variabilité morphosyntaxique des termes, permet aussi (dans une perspective contrastive) de prendre la mesure des désaccords entre les langues sur la dénomination des concepts. Cette application contrastive de l'approche textuelle est au cœur du programme de recherche sur la métalangue grammaticale évoqué plus haut.

Toutefois, si le développement de ce programme s'inscrit dans le cadre d'une telle problématique, les travaux réalisés n'en adoptent pas pour autant toute la méthodologie, dans la mesure où les entités terminologiques ne sont pas construites par extraction à partir des textes mais sélectionnées préalablement dans la langue source, puis étudiées en *corpus* dans leur diversité discursive, pour être enfin traduites par contextes.

Cela appelle une distinction précisément *terminologique* : la *terminologie textuelle* est une approche constructionnelle et élaborative des termes qui se réalise sur la base d'une analyse des discours de spécialité réunis en *corpus* ; l'*approche textuelle de la terminologie* est une étude contextuelle des termes qui sont requis dans telle ou telle activité scientifique ou technique. Dans le premier cas, il s'agit d'une analyse spécifiquement terminologique, qui est de nature à indexer les unités du vocabulaire spécialisé d'une activité, quitte éventuellement à modifier ce vocabulaire s'il en existe une indexation antérieure élaborée de manière moins empirique. Dans le second cas, il s'agit d'une analyse terminographique, qui vise à décrire la forme et le sens des termes d'une activité scientifique ou technique dans divers contextes.

3. CORPUS ALIGNÉS EN LINGUISTIQUE : LA PROBLÉMATIQUE TRADUCTIONNELLE

3.1. L'alignement

L'alignement met en correspondance, sur un support électronique, deux segments textuels (ou *contextes*) de langue différente, dont l'un est la traduction de l'autre. La correspondance des deux textes alignés s'établit à différents niveaux de structuration de l'énoncé : au niveau du mot, du syntagme, de la phrase, du paragraphe, de la section, etc., selon le type de requête. La granularité de la correspondance peut donc être plus ou moins fine.

Le recours aux textes alignés offre une réelle alternative à l'inévitable incomplétude des dictionnaires puisqu'il permet à un traducteur, qui formule une requête sur les occurrences d'un mot ou d'une expression dans un *corpus*, d'accéder à des contextes variés (concordances), et donc de trouver des solutions qu'un dictionnaire ne lui aurait pas fournies. Cette aide à la traduction est d'autant plus précieuse que les séquences traduites de la langue cible ne sont bien sûr pas des transpositions mot à mot des séquences de la langue source. Ainsi, une expression idiomatique peut faire l'objet de traductions distinctes adaptées aux différents contextes de ses occurrences. On est donc ici très éloigné du simple transcodage de la traduction automatique, et du transcodage terminologique des glossaires bilingues. Les concordances permettent de traiter les métatermes comme des expressions, et elles permettent également, dans certains cas, par ce travail de contextualisation, au-delà des notions, d'accéder aux concepts, dans le respect de la tradition grammaticale dont ils sont issus.

On exposera ici quelques-uns des problèmes rencontrés dans l'élaboration des contextes alignés, en partant de quelques exemples :

français → arabe

Polylexicalité / تعدد العجمة

Les mots composés non soudés forment des unités *polylexicales*. Cette *polylexicalité* est marquée par le fait que des constituants de la séquence voient certaines de leurs propriétés syntaxiques et sémantiques neutralisées, en particulier la capacité de faire l'objet d'une caractérisation ou d'une détermination (ex. **une pomme de bonne terre*, **un permis de bien conduire*). [F. Neveu, 2000, *Lexique des notions linguistiques*]

تكوّن الكلمات المركبة غير الملتحمة وحدات متعدّدة العجمة. و يوسم تعدّد العجمة هذا من خلال إلغاء بعض الخصائص التركيبية و الدلالية لمكوّنات الوصلة، و خاصّة القدرة على أن تكون [هذه المكوّنات] موسومة أو محدّدة (مثال : *سلاح أبيض ناصع، *رخصة سياقة جيّدة).

.trad .B .Ouerhani

وصف معرف / Description définie

<p>Depuis les travaux de Bertrand Russell, au début du XX^e siècle, on appelle description définie une expression nominale pourvue d'un déterminant défini (article ou adjectif possessif), autrement dit un syntagme nominal, expansé ou non par un modificateur (adjectif, syntagme prépositionnel, relative, etc.) : ex. la voiture, la voiture noire, la voiture de Marie, la voiture qui est garée derrière l'église, etc. [Neveu, 2000].</p>	<p>منذ أعمال برتراند روسل في مطلع القرن العشرين، نسمي وصف معرف العبارات الاسمية المحتوية على محدد معرف (التعريف بـ"ال" أو صفة ملكية)، هو بعبارة أخرى مركب اسمي يوسعه محور (صفة، مركب حرفي، موصول، إلخ.) أو لا. مثال : السيارة، السيارة السوداء، سيارة مريم، السيارة الواقفة وراء الجامع، إلخ.</p> <p>trad.B. Ouerhani</p>
--	---

محيط معتقدي / Univers de croyance

<p>Dans une première approche, on appellera « univers de croyance » ou « univers » l'ensemble indéfini des propositions que le locuteur, au moment où il s'exprime, tient pour vraies ou qu'il veut accréditer comme telles ». [R. Martin, 1992, <i>Pour une logique du sens</i>]</p>	<p>نسمي، في مقاربة أولى، "محيط معتقدي" أو "محيط" المجموع غير المحدد من الأقوال التي يعتبرها المتكلم، في الوقت الذي يتكلم فيه، حقا أو التي يريد أن تعتمد كذلك.</p> <p>trad. S. Mejri & T. Baccouche</p>
---	--

arabe → français

المركب الإسنادي / Syntagme prédicatif

<p>المركب الإسنادي هو مركب يتكوّن من التّواة الإسناديّة مع ما يمكن أن يتّصل بها من متمّمات و يحتلّ محلا من المحلات الوظيفيّة في الجملة. [كتاب اللغة، س 8 أساسي : ص 20]</p>	<p>Le syntagme prédicatif est un syntagme constitué d'un noyau prédicatif avec les compléments possibles, et qui occupe une fonction dans la phrase.</p> <p>[Manuel de langue, 8^e de base]</p> <p>trad. B. Ouerhani</p>
--	--

المركب شبه الإسنادي / Syntagme pro-prédicatif

<p>المركب شبه الإسنادي هو مركب اسمي يبدأ بمشتقّ عمل عمل الفعل. يتكوّن المركب شبه الإسنادي من : مشتقّ عمل عمل الفعل + متعلق به (أو أكثر). [كتاب اللغة، س 8 أساسي : ص 76]</p>	<p>Le syntagme pro-prédicatif est un syntagme nominal commençant par un dérivé qui fonctionne comme le verbe [prédicatif]. Le syntagme pro-prédicatif se compose de : un dérivé à valeur prédictive + un argument (ou plus)</p> <p>[Manuel de langue, 8^e de base]</p> <p>trad. B. Ouerhani</p>
---	---

3.2. La traduction de l'exemple

Une des premières difficultés de la mise en correspondance de textes grammaticaux de langue différente réside dans la traduction de l'exemple. Faut-il ou non traduire les exemples du texte français en arabe, ou du texte arabe en français ? La traduction littérale est bien sûr dénuée de pertinence dans la plupart des cas. Mais la traduction adaptée peut conduire à une exemplification problématique, étant donné l'absence fréquente de compatibilités et de parallélismes linguistiques, métalinguistiques et notionnels entre les deux langues et les deux cultures. La recherche d'un exemple différent dans la langue cible est une solution qui peut être retenue dans certains cas. Pour autant, ces difficultés ne remettent pas en question la nécessité et l'intérêt d'une traduction des exemples français en arabe ou des exemples arabes en français, ou plus largement d'une adaptation de ces exemples à la langue cible.

Eu égard à la spécificité des problèmes traductionnels posés par l'exemplification, problèmes qui varient en fonction des unités terminographiques et des sources, il ne semble pas pertinent dans ce domaine de systématiser la méthodologie. L'important est que l'utilisateur du *corpus* aligné ait accès, pour les traductions délicates, aux informations linguistiques (commentaires, choix d'exemples de natures différentes, etc.) résultant des options retenues par les traducteurs pour le contexte en langue arabe (français → arabe) ou en langue française (arabe → français). Pour ce faire, il est indispensable de retenir le principe d'un document hypertextuel, permettant, grâce à des liens créés dans certains contextes particulièrement difficiles à transposer, d'accéder à ces informations dans une fenêtre autonome. Le *corpus* aligné s'enrichit ainsi d'un ensemble de descriptions linguistiques sur les problèmes conjoints de la terminologie et de la traduction. Mais l'occasion est également donnée de concevoir toute la valeur argumentative de l'exemple en linguistique (que celui-ci soit construit ou non). En tant qu'il fournit un ensemble de données observables servant à tester la valeur d'une proposition théorique (explicite ou non), l'exemple fonctionne comme le sujet employé en mention de cette proposition. À cet égard, il illustre moins une règle qu'il ne définit une instance de réfutabilité. En cela, il ne saurait être tenu pour étranger au texte linguistique. La difficulté de transposition rendue saillante par l'alignement des contextes résulte à la fois de la confrontation des langues et de celle des métalangues, métalangues qui importent, avec leurs unités terminologiques, l'ensemble discursif où ces unités ont pris naissance. Les traductions réalisées dans le cadre de ce programme ont fait apparaître des difficultés, connues et donc prévisibles, liées à la non-coïncidence des métalangues grammaticales française et arabe : métatermes français sans correspondant en arabe (et *vice versa*), métatermes français connaissant plusieurs correspondants en arabe (et *vice versa*), difficultés encore accrues par les problèmes traductionnels liés aux abréviations, nombreuses en français mais inexistantes en langue arabe.

3.3. La sélection des contextes

Les contextes sélectionnés pour un tel travail d'alignement doivent présenter une diversité suffisante afin de fournir une base de données représentative des différents usages de la terminologie linguistique. Cela implique de sélectionner des contextes empruntés à des sources très différentes : non seulement des articles ou des ouvrages spécialisés,

mais aussi des ouvrages de vulgarisation (dictionnaires, glossaires, grammaires), qui peuvent se révéler approximatifs dans l'utilisation de la métalangue. Cette diversité garantit l'ouverture épistémologique et théorique du *corpus* (donc des contextes), *corpus* qui ne saurait être l'illustration d'une doxa ou d'un modèle, c'est-à-dire d'une pensée linguistique unique.

Un des principaux obstacles à l'harmonisation des contextes tient bien sûr à la diversité des métatermes, qui peuvent offrir des degrés de spécialisation linguistique très variables, représentatifs des différences de traitement entre les unités (p. ex. *abduction*, *anaphore associative*, *déduction*, *hyperlangue*, *induction*, *monosème*, *opérateur*, *partitif*, etc.). Il est donc nécessaire de puiser non seulement à différentes sources conceptuelles, mais aussi à différentes sources discursives et typologiques, et, de manière minimale, de distinguer entre les contextes définitionnels, fréquemment polyphoniques voire citationnels, et les contextes non définitionnels, applicatifs, qui permettent fréquemment d'observer certains flottements notionnels liés à leur mode de textualisation.

Exemples de contextes définitionnels en français

Un *monosème*, ou un mot monosémique, est un signe linguistique dont le sémantisme plénier est réalisé du premier coup, sans qu'aucune saisie précoce ni intermédiaire soit pratiquée sur la ligne du « temps opératif » utilisé pour sa formation, et où, par conséquent, le signifié de puissance se confond avec le signifié d'effet.

J. Picoche, 1986, *Structures sémantiques du lexique français*.

Un *opérateur* n'est autre que le nom d'une fonction, c'est-à-dire d'un procédé permettant de mettre en correspondance les éléments de deux ensembles. L'ensemble d'entrée s'appelle le « domaine d'application » de l'opérateur. Ses éléments sont dits « opérands » ou « arguments de l'opérateur ». Les images associées aux opérands dans l'ensemble de sortie s'appellent « résultats ».

A. Berrendonner et alii, 1983, *Principes de grammaire polylectale*.

[...] l'espace-temps, par rapport à l'intercommunication humaine, n'est pas vide, il dispose d'une certaine structure que lui confèrent les objets et les sujets qui l'occupent. Appelons *hyperlangue* cet espace-temps ainsi structuré. Introduire un nouvel objet (par ex. un sujet doté de capacités linguistiques déviantes, un dictionnaire, ou encore le moyen de communiquer à distance) change la structure de l'hyperlangue.

S. Auroux, 1998, *La Raison, le Langage et les Normes*.

Exemples de contextes non définitionnels en français

Le latin n'a pas d'article partitif ; mais il exprime la notion de *partitif* par divers tours, principalement le génitif. Pour exprimer « la moitié du butin », il dit *dimidium praedae*. Mais le génitif est parfois remplacé par un tour prépositionnel, avec la prépositionnel *de* (+ ablatif) ; Plaute, *Ps.* 1164 : *dimidium de praeda*. Formellement, le partitif français résulte de l'adjonction de l'article défini (issu de *ille*) à cette préposition *de* (d'où *du*, *de la*, *des*).

G. Serbat, 1980, *Les Structures du latin*.

La résolution des *anaphores associatives* est l'un des problèmes auxquels se trouvent confrontés les systèmes d'extraction automatique de connaissances et de simulation du raisonnement « naturel ».

D. Apothéloz, 1995, *Rôle et fonctionnement de l'anaphore dans la dynamique textuelle*.

Déduction : tous les haricots de ce sac sont blancs, ces haricots proviennent de ce sac, donc ils sont blancs. *Induction* : ces haricots proviennent de ce sac, ces haricots sont blancs, donc tous les haricots de ce sac sont blancs. *Abduction* : tous les haricots de ce sac sont blancs, ces haricots sont blancs, donc ces haricots proviennent de ce sac.

S. Auroux, 1998, *La Raison, le Langage et les Normes*.

Une déontologie scrupuleuse doit donc être respectée dans la sélection des contextes regroupés autour d'une entrée. Elle s'applique : (i) au nombre des séquences retenues, (ii) à la diversité des auteurs et des sources, (iii) à la diversité des discours linguistiques (spécialisation/vulgarisation, définition/description), (iv) à l'intérêt linguistique des contextes (on peut tolérer l'usage approximatif d'un métaterme, caractéristique d'un discours de vulgarisation, mais on doit exclure les emplois confus et erronés), (v) à la cohérence notionnelle et à l'évolution de la réflexion linguistique (le groupement doit rendre compte autant possible des éventuels changements théoriques et méthodologiques enregistrés dans l'usage du métaterme au cours du XX^e siècle).

3.4. Fonctionnement morphosémantique des termes linguistiques

Ce travail de sélection de contextes, de traduction et d'alignement, tel qu'il a été réalisé dans le cadre du programme évoqué plus haut, a fait surgir certains faits intéressants, notamment l'usage croissant des termes complexes dans le discours linguistique, usage qui témoigne d'une évolution manifeste de la discipline (voir Neveu 2005). Ce recours fréquent à la polylexicalité dans l'analyse linguistique contemporaine s'observe notamment dans la concurrence parfois mal justifiée qui s'établit entre un terme simple, disponible, fréquemment entériné dans une tradition grammaticale, et un terme complexe renvoyant à une même réalité notionnelle (p. ex. *apposition/construction apposée* ; *vocatif/segment vocatif* ; *apostrophe/terme d'adresse*, etc.). Mais au-delà de cette concurrence, le phénomène reste massif (p. ex. *description définie*, *désignateur rigide*, *univers de croyance*, *verbe support*, etc.) et témoigne de la vigueur terminologique de la science du langage.

Ce point révélé par l'étude des *corpus* est important pour caractériser la spécificité de la terminologie linguistique. Le recours à l'unité polylexicale entraîne une démultiplication des traits de nomination dans le métaterme. Mais ce libellé de l'étiquette terminologique témoigne aussi d'une approche résolument empiriste de l'activité linguistique. C'est donc tout à la fois une dénomination complexe servant à la description d'une réalité complexe et l'affichage d'une épistémologie ou d'une pensée de la langue et du langage.

Concernant la terminologie linguistique, on en vient ainsi à passer de l'idée d'un technolécite sans origine ni destination, à l'idée d'un discours orienté, pris en charge, comme discours, par un interdiscours.

4. CONCLUSION

L'approche textuelle de la terminologie linguistique, outre la constitution des bases de données organisées en *corpus* nécessaires à l'alignement des segments textuels, permet donc au terminographe d'élaborer une description des termes à partir de leur

mode de textualisation. Dans cette perspective, il ne s'agit pas tant de définir ces mots et ces expressions que de préciser les conditions dans lesquelles ils peuvent recevoir des interprétations spécifiques. Pour ce faire, un recours à la source des termes, à leur forme, à la comparaison des emplois est indispensable. Le terminographe, en sciences du langage, est nécessairement un épistémologue témoignant par l'étude du vocabulaire d'un état passé et présent de son domaine scientifique.

BIBLIOGRAPHIE

- Auroux (S.), 1989 : « Le langage et la science : une visée historique », dans Reichler-Béguelin (M.-J.), éd., *Perspectives méthodologiques et épistémologiques dans les sciences du langage*, actes du colloque de Fribourg, 1988, Berne, Peter Lang, pp. 51-68.
- Baccouche (T.) et Mejri (S.), 2000 : « Terminologie et traduction (cas de l'ornithologie et de la linguistique) », *Méta*, XLV, 3, pp. 437-444.
- Béjoint (H.) et Thoiron (Ph.), dir., 2000 : *Le Sens en terminologie*, Lyon, Presses universitaires de Lyon.
- Berrendonner (A.), 1982 : *L'Éternel Grammairien. Étude du discours normatif*, Berne, Peter Lang.
- Boisson (C.) et Thoiron (Ph.), dir., 1997 : *Autour de la dénomination*, Lyon, Presses universitaires de Lyon.
- Bourdieu (P.), 2001 : *Science de la science et réflexivité*, Paris, Raisons d'agir.
- Cabré (M. T.), 1998 : *La Terminologie – Théorie, Méthode et Applications*, Ottawa/Paris, Presses universitaires d'Ottawa/Armand Colin.
- Colombat (B.) et Lazcano (E.), dir., 1998, 2000 : *Corpus représentatif des grammaires et des traditions linguistiques*, Paris : Société d'histoire et d'épistémologie des sciences du langage (SHESL), Presses universitaires de Vincennes, hors-série n° 2 (1998), hors-série n° 3 (2000).
- Colombat (B.) et Savelli (M.), dir., 2001 : *Métalangage et terminologie linguistique*, Leuven, Paris, Peeters.
- Colombat (B.), 2004 : *Corpus des textes linguistiques fondamentaux*, « Présentation du corpus des grammaires », « Pourquoi un corpus représentatif des grammaires et des traditions linguistiques ? », < <http://www.ens-lsh.fr/labo/ctlf/> > (pages « Documents de référence »).
- Condamines (A.), 2005 : « Linguistique de corpus et terminologie », dans *Langages*, 157, pp. 36-47.
- Depecker (L.), dir., 2005 : « La terminologie : nature et enjeux », dans *Langages*, 157.
- Grossmann (F.) et Rink (F.), 2004 : « La surénonciation comme norme du genre : l'exemple de l'article de recherche et du dictionnaire en linguistique », dans *Langages*, 156, pp. 34-50.
- Hymes (D.), dir. 1974 : *Studies in the History of Linguistics. Traditions and paradigms*, Bloomington/London, Indiana University Press.
- Hymes (D.), 1983 : « Traditions and Paradigms », dans Hymes (D.), édité., *Essays in the History of Linguistic Anthropology*, Amsterdam, John Benjamins, pp. 345-383.
- Lemaréchal (A.), dir., 1999 : *La Terminologie linguistique*, Mémoires de la Société de linguistique de Paris, nouvelle série, tome VI, Paris, Peeters.
- Lerat (P.), 1995 : *Les Langues spécialisées*, Paris, PUF.
- Lerat (P.), à paraître en 2006 : « Quelques réflexions sur le traitement terminographique unilingue de la métalangue grammaticale en français », dans Neveu (F.), *La Terminologie linguistique – Problèmes traductionnels, Conceptuels, Épistémologiques, Syntaxe et Sémantique*, Caen, Presses universitaires de Caen.
- L'Homme (M.-C.), 2004 : *La Terminologie : principes et techniques*, Montréal, Presses de l'Universitaires de Montréal.

- Mejri (S.), 2003 : « Traduction linguistique : problème terminologique ou construction conceptuelle », dans Mejri (S.), Gross (G.), Baccouche (T.), Clas (A.), dir., *Traduire la langue, Traduire la culture*, Paris, Tunis, Maisonneuve et Larose, pp. 177-191.
- Milner (J.-C.), 1989 : *Introduction à une science du langage*, Paris, Le Seuil.
- Neveu (F.), 2004 : *Dictionnaire des sciences du langage*, Paris, Armand Colin.
- Neveu (F.), 2005 : « Sur l'usage des termes complexes dans le discours de la science du langage. Préliminaire à une étude comparée de la terminologie linguistique », dans Mejri (S.) et Thoiron (Ph.), *La Terminologie, entre traduction et bilinguisme*, Tunis, AUF, réseau LTT, pp. 107-120.
- Neveu (F.), à paraître en 2006 : « La notion de tradition grammaticale et son usage en linguistique française », *Langages*, dans Mejri (S.) et Neveu (F.), édit., « La tradition grammaticale ».
- Neveu (F.), dir., à paraître en 2006 : « La Terminologie linguistique – Problèmes traductionnels, conceptuels, épistémologiques », *Syntaxe & Sémantique*, Caen, Presses universitaires de Caen.
- Rastier (F.), 1998 : « Le problème épistémologique du contexte et le statut de l'interprétation dans les sciences du langage », *Langages*, 129, pp. 97-111.
- Slodzian (M.), 2000 : « L'émergence d'une terminologie textuelle et le retour du sens », dans Béjoint (H.) et Thoiron (Ph.), *Le Sens en terminologie*, Lyon, Presses universitaires de Lyon, pp. 61-85.
- Swiggers (P.), 1999 : « Pour une systématique de la terminologie linguistique : considérations historiographiques, méthodologiques et épistémologiques », dans Lemaréchal (A.), édit., *La Terminologie linguistique*, Mémoires de la Société de linguistique de Paris, nouvelle série, tome VI, Paris, Peeters, pp. 11-49.
- Thoiron (Ph.), dir., 1996 : « La dénomination », *Méta*, 41/4.

LES NOMS DE DIVINITÉS : WEB, CONTEXTES ET CLASSES D'OBJETS

Michel Mathieu-Colas

laboratoire de linguistique informatique (LLI)

CNRS et Université Paris XIII

Villetaneuse, France

Résumé : *Cette étude porte sur un exemple de classe d'objets : les noms de divinités. Nous appuyant sur un premier travail de description de la classe (dictionnaire électronique et analyse syntaxique), nous nous interrogeons sur les nouvelles possibilités offertes par le Web ; après avoir constitué un large corpus, nous étudions quelques exemples de concordances. Le résultat de cette expérience est double : on ne peut tout attendre de l'accumulation de données brutes, mais le Web, adossé à une analyse linguistique, peut contribuer efficacement à la description.*

Mots-clés : *classes d'objets, noms de divinités, corpus, Web*

Nous voudrions rendre compte ici d'une expérience qui met en jeu, tout à la fois, la notion de « classe d'objets », telle que nous l'avons élaborée au LLI, et les nouvelles ressources disponibles sur le Web. Nous partirons, à titre d'exemple, de la classe des noms de divinités, que nous avons déjà décrite, il y a une dizaine d'années, avec des procédés classiques (dictionnaires, encyclopédies, ouvrages spécialisés pour la constitution de la nomenclature, et sélection de textes pour la description de la syntaxe). Il s'agira de mettre cette description à l'épreuve du Web et de tester, en même temps, ce que le Web peut apporter ou non à ce type d'étude ¹.

Plus précisément, nous proposons la démarche suivante :

- description de la classe ;
- constitution d'un *corpus* extrait du Web ;
- élaboration de concordances pour étudier l'environnement des noms de divinités. Il conviendra d'examiner si le Web confirme les analyses antérieures et s'il permet de les affiner ou de les compléter ;
- utilisation du contexte pour enrichir la liste des noms. Cette expérience est complémentaire de la précédente puisqu'il s'agit cette fois de déterminer si la description syntaxique, projetée sur le Web, permet d'identifier de nouvelles unités.

¹ Je tiens à remercier particulièrement Marc Chanove pour l'aide précieuse qu'il m'a apportée dans la réalisation de ce travail, sur le plan informatique.

1. PRÉSENTATION DE LA CLASSE

1.1. Rappel théorique

Rappelons brièvement ce que nous entendons par « classes d'objets »². Ce sont des ensembles d'unités lexicales offrant une double homogénéité, syntaxique et sémantique, les deux aspects se trouvant, selon nous, étroitement corrélés (on pourrait tout aussi bien parler de classes « sémantaxiques »). C'est ainsi, par exemple, que les noms désignant des voies de communication (des mots comme route, rue, chemin, sentier, allée, etc.) offrent à la fois une unité sémantique – ils désignent des espaces aménagés pour aller d'un lieu à un autre – et une cohérence syntaxique : ils entretiennent une relation privilégiée, dans le cadre des phrases, avec un certain nombre de prédicats appropriés : des verbes (emprunter), des adjectifs (impraticable), des prédicats nominaux (aménagement), etc. Cette mise en relation du sens et du contexte s'avère opératoire, en particulier, pour le traitement de la polysémie et la réduction des ambiguïtés : on n'emprunte pas une route comme on emprunte de l'argent³...

1.2. La classe des noms de divinités

S'agissant des noms de divinités, l'étude (inédite) que nous avons entreprise comportait deux volets :

a) Un dictionnaire électronique comportant plus d'un millier d'entrées, avec, pour chaque nom, un certain nombre d'informations complémentaires : domaine géographique (Grèce antique, Inde, Afrique), signification et attributions (dieu de la mer/déesse de la moisson), variantes (graphiques ou phonétiques : *Anâhitâ*, *Anâhita*, *Anaitis*, *Anahitis*, *Anahid*), sources d'information (dictionnaires et encyclopédies, traités de mythologie ou d'histoire des religions, études spécialisées).

Voici, à titre d'exemple, les douze premières entrées du dictionnaire :

Ndiv	GENRE	DOMAINE	SOUS-DOM	SENS	VAR	SOURCES
A-a	n.f.	Mésopotamie	assyro-babyl.	épouse du dieu-soleil	Aia	GDEL ; LXX ; PLE1 (: 196)
Abéona	n.f.	Romains		divinité des voyages		SCH
Abo	n.m.	Afrique	Kaffa	Être suprême (solaire)		ELIA (:119)

² Cette notion, que nous avons développée avec Gaston Gross à partir du début des années 90, a fait l'objet, entre autres, d'un numéro de *Langages* (voir la bibliographie). On pourra s'y reporter pour une présentation d'ensemble.

³ Pour l'intérêt des classes d'objets dans le cadre du traitement de la polysémie, voir ici même la contribution de Salah Mejri.

Acca Larentia	n.f.	Romains		divinité de la fertilité des champs	Larentia	GDEL ; SCH ; GRI
Achéloos	n.m.	Grecs		dieu-fleuve	Achéloüs	EL (: 536)
Açvins	n.m. pl.	Inde	véd.	dieux jumeaux de la lumière	Ashvins, Asvin	PAY (:196) ; QSJ-H (:10) ; Enc. Bordas (:141);LXX ; ELIA (:91)
Adad	n.m.	Mésopotamie	assyro-babyl.	dieu de l'orage	Hadad	GDEL ; QSJ-PO (:43) ; PR2
Addou	n.m.	Syrie, Canaan	Amorrites	dieu de l'orage		GDEL (Baal) ; PLE1 (: 316)
Adéona	n.f.	Romains		divinité des voyages		SCH
Aditi	n.f.	Inde	véd.	mère des dieux		GDEL (Aditya) ; QSJ-H (: 9) ; Encyclo. Bordas (:141) ; ELIA (: 81)
Adonis	n.m.	Phénicie		dieu de la végétation	Adônis	GDEL ; THIO ; ELIA (: 136)
Aegipan	n.m.	Grecs		divinité champêtre (surnom de Pan)	Egipan	GDEL ; LXX

b) Une étude syntaxique décrivant les prédicats appropriés et plus généralement toutes les collocations caractéristiques de la classe. Nous en donnons seulement quelques exemples :

– Verbes

Ndiv en position de sujet :

Ndiv (*créer + engendrer + faire*) N (le monde, la Terre, les hommes, etc.)

Ndiv *protéger* N (la végétation, les hommes, etc.)

Ndiv *juger* Nhum (les hommes, les morts)

Ndiv *apparaître* (à Nhum) (*sous la forme de N*)

Ndiv en position d'objet :

Nhum *adorer* Ndiv

Nhum *rendre un culte à* Ndiv

Nhum *offrir un sacrifice à* Ndiv

Nhum (*célébrer + participer à*) *les mystères de* Ndiv

N-hum (animal, etc.) *représenter* Ndiv

– **Adjectifs**

Ndiv être Adj-qual (Adj-qual = : *bienfaisant, miséricordieux, secourable, favorable, bon, juste*, etc.)

Ndiv être Adj-défaut (Adj-défaut = : *malfaisant, redoutable, terrible, cruel*, etc.)

– **Noms prédicatifs**

Nhum être un *adorateur de* Ndiv

N être un *messenger de* Ndiv

Ndiv être *le maître de* (le Ciel, la Terre, les eaux, etc.)

Ndiv être *l'incarnation de* Nabstrait (la Loi, etc.)

Ndiv être (*le fils + la fille*) de Ndiv

– **Collocations diverses**

La statue de Ndiv

Le mythe de Ndiv

Le sanctuaire de Ndiv

(*devant + en présence de*) Ndiv

par Ndiv ! (*par Jupiter, par Toutatis* !)

En tout, nous avons mis en évidence 220 descripteurs. Ceux-ci ont été établis par l'analyse de textes relatifs aux différentes mythologies. Nous citerons notamment le « Que sais-je ? » relatif aux *Dieux de l'Égypte* (Daumas 1977), que nous avons analysé dans le détail. On y trouve naturellement, à chaque page, des noms de divinités insérés dans des phrases : « À Héliopolis, ATOUM *revêtait la forme humaine*, au moins à l'époque historique. À Atfih, HATHOR, *déesse de l'amour, avait figure de femme*, mais laissait passer de sa perruque des oreilles de vache. Héracléopolis *rendait un culte au dieu-bélier HARSAPHÈS*. THOT, *à tête d'ibis, était patron d'Hermopolis*. À Siout, OPHOÏS *se montrait sous l'apparence d'un chacal*. HORUS d'Edfou *avait pour animal sacré un faucon dont les imagiers adaptaient la tête à son corps humain*. KHNOUM, à Esna ou à Éléphantine, *présentait une tête de bélier* (p. 17 ; nous soulignons). »

Le relevé minutieux des contextes d'occurrence nous a permis de constater, au fil des pages, une saturation progressive de la description : la poursuite de la lecture apportait de moins en moins de nouveautés syntaxiques, ce qui tendrait à prouver, sinon l'exhaustivité de notre étude, du moins la cohérence de l'ouvrage.

Tels sont les principaux aspects de notre première exploration. Il s'agit à présent de la confronter au Web.

2. CONSTITUTION DU *CORPUS*

Puisque nous souhaitons trouver des pages où il est question de divinités, en vue de disposer d'un large éventail de contextes, le plus simple est de partir de notre dictionnaire. Nous avons donc procédé aux opérations suivantes.

2.1. Sélection de noms de divinités

Pour faciliter le traitement, nous avons omis certains types d'unités :

- les noms relatifs aux religions monothéistes (*Yahvé, Christ, Allah*), dont la fréquence sur le Web eût écrasé les autres dénominations ;
- les noms composés (ex : *Bona Dea*, déesse romaine de la fécondité ; *Cao Dai*, Être suprême vietnamien) ;
- les pluriels (*les Euménides, les Pléiades*) ;
- les noms trop ambigus (le maintien d'appellations comme *Bel* ou mieux encore *Eros* aurait induit un trop grand nombre de pages non désirées...).

Au terme de cette sélection, nous avons retenu 859 noms, puisés dans le dictionnaire électronique.

2.2. Recherche d'adresses URL

Cette liste a été introduite sous *Google* (avec une automatisation partielle de la recherche). Mais il est vite apparu que le résultat comportait trop de « bruit » et trop d'ambiguïtés : de nombreux noms de divinités peuvent être utilisés comme des noms ou des prénoms humains, voire des noms de marques, des désignations d'agences de voyages, etc. Afin de limiter les parasites, il a été demandé que, dans les pages recherchées, figure au moins un des lexèmes *dieu, déesse* ou *divinité*.

D'autre part, pour les noms les plus fréquents, nous nous sommes limités aux 800 premières réponses. En tout, 59 000 adresses ont été identifiées (soit une moyenne d'environ 70 pages par nom, avec bien entendu des écarts considérables : 80 000 pages pour Jupiter, 2 seulement pour Karagabi, un dieu céleste de Colombie).

2.3. Récupération et filtrage des pages

Les programmes mis au point par les informaticiens du LLI ont permis de récupérer (d'« aspirer ») la plupart de ces pages. Si l'on met de côté 7 500 erreurs d'adresse et 5 000 fichiers *Word* ou *Acrobat Reader* (.doc et .pdf) non traitables directement, le nombre de pages HTML récupérées s'élève à plus de 46 500 (plus de 1,8 Go). Il a fallu ensuite procéder à un filtrage – réduction des balises HTML et de toutes sortes de signes parasites – afin d'obtenir textes aussi « normaux » que possible (nous avons dû toutefois renoncer provisoirement à l'apostrophe).

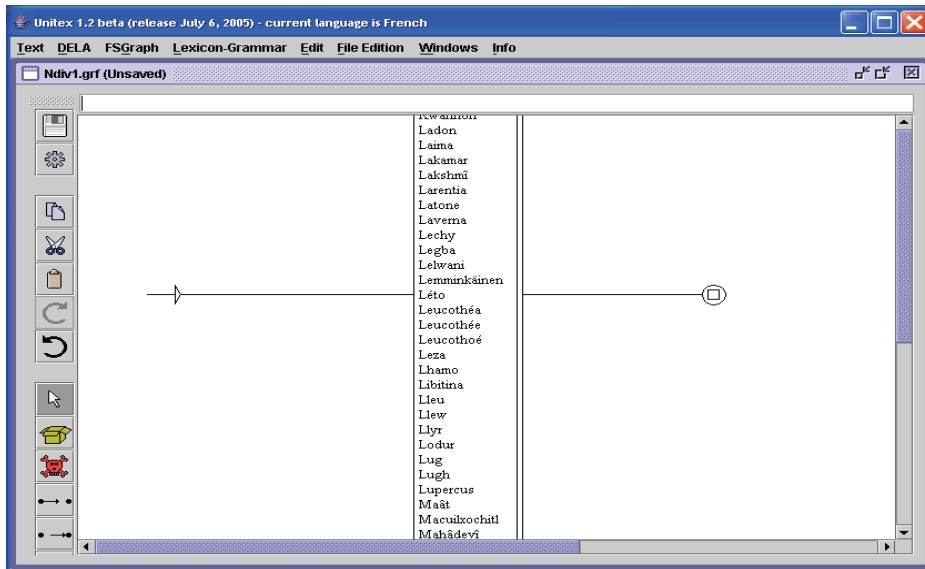
2.4. Élaboration de concordances

La dernière étape informatique a consisté à intégrer ces pages dans *Unitex*. Il s'agit d'un programme mis au point à l'Université de Marne-la-Vallée (Paumier 2004), et qui autorise l'élaboration de concordances à partir d'expressions régulières et de graphes. Une de ses particularités, et non la moindre, est d'intégrer des dictionnaires, ce qui permet d'insérer dans les expressions recherchées des catégories morphosyntaxiques (nom, verbe, etc.). Pour des raisons techniques, l'importation sous *Unitex* n'a été possible que pour une partie seulement des pages récupérées (env. 20 500). Certains résultats ont par la suite été transférés sous *MS Access* afin d'en faciliter le traitement.

3. ÉTUDES DE CONCORDANCES

3.1. Le contexte des noms de divinités

Pour obtenir, sous *Unitex*, une concordance des noms de divinités, il suffit d'intégrer les 859 noms de notre liste au sein d'un graphe élémentaire :



À partir de ce graphe, le programme produit la liste des contextes :

e Adéona ceux qui nous abordent, à la déesse Abéona ceux qui s'en vont {S} l ? pourquoi fallait-il s
ona {S} Divinités qui président au voyage. {S} Abéona est invoquée par ceux qui partent abire, Adéona
lta de {S} Nestos, Mesologii embouchure de l Achéloos, les régions de montagne Achaia, Arkadia,
ui disputa à Hercule la main de Déjanire. {S} Achéloos ayant la faculté de se transformer, Ovide, dan
ui arracha une de ses cornes. {S} En échange, Achéloos dut lui donner une corne de la chèvre Amalthée,
la chèvre Amalthée, la nourrice de Zeus. {S} Achéloos est le père de plusieurs nymphes : Callirhoé,

Ses {S}Enfants {S}Achéloos {S}Dieu du fleuve Achéloos les {S}Sirènes {S}Aither {S}La {S}Lumière de }Il était le fils d'Océan et de {S}Téthys. {S}Achéloos lutta avec Héraclès pour la main de Déjanire, m ours d'Océan et de Téthys 1, {S}2, {S}3. {S}Achéloos, dans son combat avec Heraklès, se transforme {S}Le plus grand des fleuves de la Grèce, l Achéloos, est le plus vénéré. {S}Les dieux-fleuves sont fille du premier roi d'Athènes {S}Eréctée, Achéloos, le dieu-fleuve, est près du ruisseau, et les rer à lui. {S} Son combat avec le dieu-fleuve Achéloos, qui convoite la même jeune fille que lui _ Dé s. {S} Elle est consacrée aux Nymphes et à {S}Achéloos, si l on en juge par ces figurines et par ces se, alors qu'elle se baignait dans le fleuve Achéloos. {S} Hermès offrit de lui rendre la sandale en Les Cyclopes Les Fleuves _ Les Naïades _ L Achéloüs _ Alphée et Aréthuse L Eurotas, le Pamise, la es mythes d'Augé et de Téléphe; {S} enfin sur Achéloüs à qui il enlève Déjanire, la fille de Oeneus, adorés par leurs riverains, mais que seul l Achéloüs était tenu pour un dieu par tous les hommes 2 rincipal fleuve de la Grèce centrale, de cet Achéloüs que l on considérait cumule le fleuve-type, le ime, qui prit dès lors le nom d'Alphée. {S}L Achéloüs s'appelait d'abord le Thestius : "Voici à quel ison du volume de ses eaux, les Étoliens à l Achéloüs, en raison des traditions qui le concernent; {S} elle occasion il reçut le nom d'Achéloüs. {S}Achéloüs, fils d'Arès et de la nymphe Naïs, ayant eu co

Après élimination des doublons, nous avons obtenu près de 280 000 lignes de concordance (soit autant de contextes attestés des noms de divinités).

L'examen de ces contextes permet d'effectuer, à grande échelle, deux opérations décisives pour notre étude : d'une part l'illustration et la validation de schémas déjà connus, d'autre part la recherche de nouveaux schémas.

a) *Fonction d'illustration*

Parmi les constructions appropriées aux noms de divinités, nous avons identifié la séquence : *offrir* Dét *sacrifice* à Ndiv. Notre *corpus* fournit immédiatement des exemples de cette structure :

contexte G	Ndiv
lon offrait des sacrifices à Héphaïstos et à	Aphrodite
se, j offrirai un sacrifice à ton grand dieu	Apollon
r offrir pieusement un sacrifice à la chaste	Diane
ant d offrir des coqs en sacrifice à Eshmoun-	Esculape
ctions de Médée, il offrit un sacrifice à	Hécate
. On offre des sacrifices à Dionysos et à	Pan

contexte G	Ndiv
s offrant des sacrifices à Fébruus avatar de	Pluton
roi avait refusé d offrir en sacrifice à	Poséidon
alia, où l on offre un sacrifice à la déesse	Tacita
i les offre en sacrifice à Héra, la femme de	Zeus
S} Celui-ci voulait offrir un sacrifice à	Zeus
rrêtée, il débarque et offre un sacrifice	Zeus

Chacun de nos 220 descripteurs peut être ainsi confronté au Web, dans un but de validation et d'exemplification.

b) Fonction d'enrichissement

Cette deuxième fonction est encore plus importante pour nous, puisqu'elle a une valeur directement heuristique. L'examen des concordances fait apparaître des structures non encore identifiées. C'est le cas, par exemple, du participe *associé* :

contexte G	Ndiv	contexte D
rtagé avec Anu. Anu est souvent associé à	Adad	dans les invocations. L Adad nommé et les div
Attis parmi les Phrygiens et les Romains et	Adonis	parmi les Grecs. Il a été associé à Mithra
s ont adopté un culte de fertilité associé à	Adonis	qui a été détruit par un sanglier, un animal s
Bhgu Crépitant, géniteur prajpati associé à	Agni	, fils de Varua et Cara, père de ukra et cka;
Dieu. Dans le Vedas le nom est associé à	Agni	, le dieu du feu, qui a eu sept langues clignotan
nfini, de l éternité et de l immuabilité.	Agni	le dieu associé à la recherche de la vérité est
e par excellence est celui de Roi, associé à	Agni	le Sacerdoce, comme la Lune et le Soleil. Il
Saint, et il est identifié avec le soleil à	Ahura	Mazda ; Asha, également associé à la lumière
Asha; en vieux perse, Arta. Le nom de	Ahura	Mazda fut parfois associé à celui de Mithra ; {S
ibex ou de bouquetins, peut-être associés à	Almaqah	, des têtes de taureau, peut-être associées au
garasu C est un oiseau mythique associé à	Amaterasu	, la déesse du soleil. Représenté par un
Shrine grand d Ise, par exemple, est sacré à	Amaterasu	. Puisqu elle est associée à la famille i
pire. On y vénère en grande pompe le dieu	Amon	, associé en triade à Mout et à Khansou. Pour
x reste à Thèbes : c est sa divinité locale,	Amon	, associée à Râ, qui devient bientôt prépondérant
, elle peut être l oeil de Ra; associée à	Amon	, elle devient l une des formes de Mout, comme le

contexte G	Ndiv	contexte D
, dans le delta du Nil. Le roi des Dieux,	Amon	, fut souvent associé à RÊ, Dieu du soleil, et la
vant pour le culte du roi divinisé associé à	Amon	, la principale divinité thébaine; les anciens
t a été fortement convoitée, par exemple par	Amon	, Monton ou Khnoum. Il sera donc associé à cha
ux associés au dieu soleil Rê, et à Ptah.	Amon	, Rê et Ptah étaient les principaux dieux de l'em
k à Thèbes. Liens de famille : Associée à	Amon	. Représentation : Femme portant la Couronne r

Les quelque 600 attestations que nous avons découvertes nous conduisent à ajouter à notre description une nouvelle construction appropriée :

N (humain, animal ou attribut) *associé* à Ndiv

Nous proposons de reprendre ce double fonctionnement (illustration/enrichissement) en étudiant plus précisément deux schèmes particulièrement caractéristiques :

Verbe Ndiv (ex : les Gaulois adoraient Grannus)

Nom de Ndiv (ex : le temple de Jupiter)

3.2. Le schème [Verbe Ndiv]

Dans notre première étude, cette structure était représentée par une quinzaine de verbes :

Nhum *adorer* Ndiv

Nhum *blasphémer* (*le nom de*) Ndiv

Nhum + Ndiscours (prière, hymne, etc.) *célébrer* Ndiv

Nhum *craindre* Ndiv

Nhum *glorifier* Ndiv

Nhum *honorer* Ndiv

Nhum *implorer* Ndiv

Nhum *invoquer* Ndiv

Nhum *prier* Ndiv

Nhum *redouter* Ndiv

Nhum *remercier* Ndiv (*pour + de*)

Nhum *révéler* Ndiv

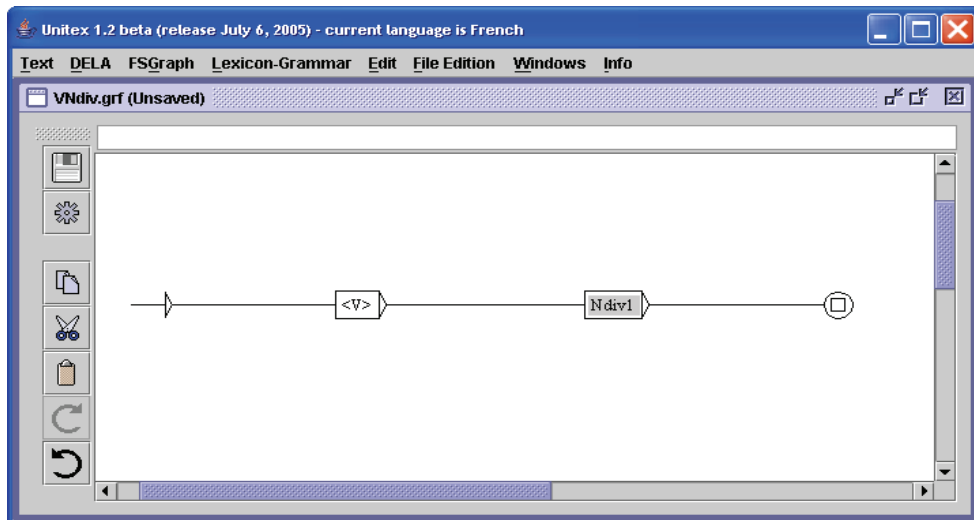
Nhum *supplier* Ndiv *de V*

Nhum *vénérer* Ndiv

N-hum *représenter* Ndiv

Nloc (temple, etc.) *abriter* Ndiv

Il est aisé de créer un graphe sous *Unitex* pour représenter cette construction, du fait que le programme sait identifier les verbes (puisque'il intègre un dictionnaire du français), et que par ailleurs on lui fournit, sous forme de sous-graphe, la liste des noms de divinités :



On obtient ainsi une liste de concordances d'environ 20 000 lignes.

Si le résultat, en termes de volume, est conforme à notre attente, il n'en va pas tout à fait de même du point de vue qualitatif. En effet, la liste des verbes les plus fréquents dans cette position est loin de correspondre à notre sélection. Voici les vingt premiers (après lemmatisation) :

Vlemme	Fréq
être	2 408
appeler	656
avoir	653
entrer	644
contrer	521
voir	435
dire	344
épouser	330
devoir	329
nommer	320

pouvoir	285
représenter	275
suivre	226
faire	222
naître	207
devenir	171
trouver	169
tuer	158
adorer	141
servir	126

Les verbes généraux se taillent la part du lion (*être* et *avoir*, bien sûr, mais aussi *voir*, *dire*, *devoir*, *faire*...), il y a des parasites dus au programme (*entre* et *contre* sont retenus par *Unitex* parce qu'ils pourraient être des formes verbales), et les verbes les plus fréquents ne sont pas nécessairement les plus spécifiques.

Cela signifie clairement qu'on ne peut se fier aux statistiques pour sélectionner les verbes appropriés, et encore moins automatiser l'opération. L'identification des opérateurs exige la compétence linguistique d'un locuteur humain.

Cette réserve étant faite, la concordance garde tout son intérêt, sur les deux plans que nous avons définis.

a) L'exemplification

C'est ainsi que le verbe *invoquer*, qui figurait dans notre première sélection, est représenté par plus de 100 attestations, dont voici quelques-unes :

contexte G	V_Ndiv	contexte D
s du ciel étaient allumées. "Nous avons	invoqué Adad	, le dieu au foudre. "Nous avons posé une q
édiction du lépreux Leper s curse En	invoquant Agni	, Indra et d autres dieux de vie, et en l
da: Avant tous les autres dieux il faut	invoquer Agni	. Nous prononcerons son nom vénérable a
se retire des sacrifices Les brâhmanes	invoquent Agni	, qui leur apparaît et leur donne la rais
t alors échouer devant la ville ; il	invoqua Amon	. La bataille s acheva par un très impor
alle du jugement. On avait coutume d	invoquer Anubis	avant toute intervention chirurgicale o
es en pierre du pays. C est là qu on	invoquait Apollon	, Boreo et Damone, divinités tutélaire

contexte G	V_Ndiv	contexte D
n pleine campagne militaire, le général	invoque Apollon	et lui promet une part du butin Tite-Li
est passée. Pour peu que je voulusse	invoquer Apollon	, J en ferais, pour vous plaire, un
eu et en leur propre nom, les Béotiens,	invoquant Apollon	et les autres divinités du temple, le
i, et bien, lauteur de la malédiction a	invoqué Apophis	, et toi tu dois entrer en contact avec
terrain de Baal pour forcer le roi qui	invoque Baal	, et lui retirer son prestige, son pouvoir,

b) La découverte de nouvelles structures

Si l'automate ne peut, à lui tout seul, construire *ex nihilo* la liste des verbes appropriés, il peut du moins offrir au linguiste un éventail de possibilités pour l'aider à développer les descriptions et à les compléter. Nous avons trouvé des verbes comme *louer*, *servir*, ou à l'inverse *défier*, *insulter*, *offenser* qui n'avaient pas été identifiés. C'est aussi le cas du verbe *fêter* (18 attestations) :

contexte G	V_Ndiv	contexte D
réunissait l ensemble des citoyens pour	fêter Athéna	. Phidias avait sans doute établi le pro
q de vin, afin que je puisse impunément	fêter Bacchus	Point de fracas ni de cris de guerre;
mes des Amnites, en face du continent,	fêtent Dionysos	. Artémidore, in Strabon, Géographie,
monter aux fêtes des anciens Grecs, qui	fêtaient Dionysos	Bacchus, dieu de la vigne et de livr
re : Indrajatra : Fin de la mousson, on	fête Indra	, le Dieu de la pluie. Octobre : Fête nati
remier jour des calendes de janvier ils	fêtaient Janus	, dieu aux deux visages. La fête symbo
s et des esclaves déguisées en matrones	fêtent Junon	Caprotina, sous le figuier; elles boive
ture ville d Athènes. Les Romains	fêtaient Neptune	tous les mois de février. texte tiré d
future ville d Athènes. Les Romains	fêtent Neptune	tous les mois de février. Son nom a é
de la lumière et du feu. Les romains	fêtaient Saturne	, dieu de l agriculture du 17 au 24
e à une époque de l année où, jadis, on	fêtait Wodan	. Cette fête a repris à son compte bon n
ythologies locales. Les Anglo-saxons	fêtaient Wotan	au solstice, les Germains par peur des f
sbos. Des jeux sportifs à Némée pour	fêter Zeus	. Des concours de théâtre et des courses a

On observera que les mythologies les plus diverses sont mises à contribution (grecque et romaine, mais aussi germanique et hindoue), ce qui témoigne de la généralité de la construction.

3.3. Le schème [Nom de Ndiv]

Nous avons procédé de façon analogue pour le deuxième schème analysé : construction du graphe et projection sur le *corpus*. Il en est résulté quelque 50 000 lignes de concordance. Les conclusions, pour l'essentiel, sont identiques aux précédentes.

a) Illustration de structures connues

Soit par exemple *le sanctuaire de Ndiv* :

contexte G	N de Ndiv	contexte D
la XIIe dynastie fut fondée à Thèbes un	sanctuaire d Amon	, lequel, agrandi successivement par l
res entourés d enceintes: au centre, le	sanctuaire d Amon	, que les Anciens considéraient comme
e, responsable des chantiers du premier	sanctuaire d Amon	. Blocs de réemploi, linteaux, stél
res entourés d enceintes: au centre, le	sanctuaire d Amon	_ de 600 m de côté _, que les Anciens
de Siouah abritait dans l Antiquité un	sanctuaire d Amon	auprès de lequel Alexandre le Grand p
t par les frémissements des dattiers du	sanctuaire d Amon	-Râ et l oracle grec délivrait ses ord
n a ravivé les couleurs. Salle Nord,	sanctuaire d Amon	-Rê-Osiris salle à gauche en arrivant
enet. Les salles du sanctuaire Le	sanctuaire d Amon	-Sokaris-Osiris salle de droite. S
. Il se divise en trois parties : le	sanctuaire d Amon	, l enceinte de Mout au Sud et celle d
lus proches d Athènes, célèbre pour son	sanctuaire d Aphaia	, ancienne divinité patronne de l il
enir les groupes de Grecs, à l abri des	sanctuaires d Aphro- dite	, Héra et Déméter, aux marges de
squ à la fin de l époque romaine. Le	sanctuaire d Aphrodite	était tellement riche qu il poss
avec la présentation de la fouille d un	sanctuaire d Aphrodite	à Théra par Ch. Sigalas.Retou
S}Cythère : île grecque de la mer Egée,	sanctuaire d Aphrodite	. Achille : v.ci-dessus. Ti
ient des marins de tous les pays. Le	sanctuaire d Aphrodite	, sur l Acrocorinthe, était renom
éloponnèse par le Cap Malée; célèbre	sanctuaire d Aphrodite	.D DéméterDéesse grecque de la

b) Enrichissement de l'analyse

Nous n'avions pas repéré, dans notre premier recensement, le syntagme *adepte de Ndiv*. Le Web le rappelle à notre attention :

contexte G	N de Ndiv	contexte D
e Lycurgue, à {S}Délos il serait devenu	adepte d Apollon	, à Delphes il aurait assisté au mystère
e de Néron, 56. {S}Othon aurait été un	adepte d Isis	: "On le vit souvent en habit de lin, com
rs le nord. {S}Mon père était un ardent	adepte d Odin	et Thor et me disait fréquemment qu'ils é
son voeu de chasteté, Daphné devient une	adepte de Diane	.{S} Arthus, laquelle la plupart des c
un abri à Vrindavan. {S} Cette nouvelle	adepte de Krishna	chanta des versets célèbres et les ve
e beau-père. {S} Mais Supatar, le prêtre	adepte de Mithra	, refusa malgré la torture qui lui fut
} : Clic ! {S} Saint Maurice était-il un	adepte de Mithra	{S}?: Clic ! {S}HELP !{S} Recherche d
mardi 14 octobre 2003 22:57 {S} Etant un	adepte de Shiva	, je proposerai à mon fils de porter à l
en forme d'épi de maïs. {S} Shivaïste :	Adepte de Shiva	, portant le trident, des lignes horizon

Ajoutons que, pour ce type de construction, la fréquence semble plus significative que dans le cas précédent. Si on regarde les « têtes de liste », on observe les fréquences suivantes :

Nlemme	Fréq
fil	4 862
temple	3 075
fil	1 781
nom	1 468
culte	1 358
épouse	727
statue	632
honneur	615
prêtre	503
frère	447
mère	436
enfant	395
sanctuaire	389
femme	357

Nlemme	Fréq
ordre	316
mythe	312
naissance	312
fête	289
forme	284
père	283
mort	276
sœur	238

On notera la surreprésentation des noms de parenté – fils, fille, épouse, frère, mère, enfant (s), femme, père, sœur : les histoires mythologiques semblent être avant tout des histoires de famille... On constate aussi la présence, en bonne place, des substantifs naissance et mort : en dehors du monothéisme, les dieux ne sont pas toujours éternels. Les autres noms qui figurent ici ne sont pas surprenants (temple, culte, prêtre, statue, etc.). Quant au substantif honneur, il ne doit sa place qu'à la locution en l'honneur de et à ses variantes (en honneur de, à l'honneur de, pour l'honneur de).

L'analyse des concordances découvertes sur le Web présente donc un intérêt certain pour la description syntaxique des classes. Une liste de noms étant donnée, elle aide au repérage des prédicats appropriés et des collocations les plus caractéristiques. Si elle ne peut se substituer au travail linguistique, elle peut du moins l'accompagner et le prolonger.

4. RECHERCHE DE NOUVELLES UNITÉS

Reste à examiner la procédure inverse. Une description du contexte étant donnée, le Web peut-il aider le linguiste à identifier de nouveaux éléments de la classe ?

Nous testerons cette possibilité sur un exemple précis. On vient de voir que, parmi les formes [Nom de Ndiv] les plus caractéristiques, figurait la séquence *culte de Ndiv*. La question est la suivante : si l'on extrait du Web tous les voisinages de *culte de*, a-t-on des chances d'identifier un pourcentage appréciable de noms de divinités et, si oui, les éléments trouvés sont-ils suffisamment riches pour compléter les listes dont nous étions partis ?

Dans la mesure où nous ne manipulons plus des classes (classes d'objets : noms de divinités ou classes grammaticales : nom, verbe, etc.), il n'est plus nécessaire, à ce

niveau, d'utiliser *Unitex*, et nous pouvons interroger directement les 46 500 pages à l'aide de filtres usuels comme les commandes *grep* ou *sed*⁴.

Le résultat se présente sous cette forme :

vertis au culte de Mithra le répandre
 iècles le culte de Mithra opposa une g
 terre. Le culte de Gu se continue ajuo
 devins du culte de Fa, le grand ancêtr
 rifier au culte de la première et pour
 issent du culte de la divinité que les
 grecs. Ce culte de Sérapis aurait duré
 Zeus. Le culte de Sérapis était très
 rieur. Le culte de l'intérieur est pra
 estiné au culte de la déesse HATHOR et
 hrygie le culte de Men, vieille divini
 agore. Ce culte de Dana a donné son no
 bre et le culte de l'arbre sacré en Cr
 ncipal du culte de la déité dans la co
 e dans le culte de Marduk et de son ép
 t pas. Le culte de Slaanesh est secret
 que "son" culte de Khaine a dévié vers
 ssocié au culte de Bacchus Dionysos. Ê
 ssocié au culte de Zeus Ammon que l on

Ayant délimité la chaîne *culte de* par des séparateurs, nous avons importé le résultat sous *MS Access* pour pouvoir le traiter avec plus de facilité. Après élimination des doublons et d'un certain nombre d'erreurs manifestes, le nombre de concordances s'est établi à plus de 5 000, réparties de façon à peu près égale entre les noms communs et les noms propres.

Les noms communs correspondent, pour une bonne part, à des termes désignant des divinités, ou directement associés à elles (*culte de l'Être suprême, culte de la Grande Mère, culte de la croix*, etc.), et pour le reste à des entités profanes, mais sacralisées par la société (*le culte de la raison, de la patrie, de la nature, de la jeunesse, de l'argent, de la consommation*, etc.).

Pour ce qui est noms propres – objet spécifique de cette recherche –, nous n'avons gardé que les noms simples, afin de pouvoir les comparer à notre liste de départ. Encore a-t-il fallu écarter un certain nombre d'« intrus », dus à des ambiguïtés syntaxiques⁵. Il nous est resté, pour finir, un peu moins de 2 400 concordances, soit 500 noms différents répondant au schéma : *culte de* Npropre.

⁴ La commande *sed*, en particulier, permet de limiter l'extension des concordances (nous avons choisi 10 caractères à gauche et 20 à droite), ce qui serait impossible avec le seul usage de *grep*.

⁵ Voir par exemple : *le culte de Dante pour Béatrice* (Dante est ici le sujet, et non l'objet du culte), *Histoire particulière et générale des religions et du culte* de Delaunaye, et surtout les emplois adjectivaux du mot *culte* : *film culte* d'Antonioni, de Coppola, de David Lynch, *roman culte* de Boulgakov ou de Roald Dahl, *concert culte* de Nailbomb, etc.

143 d'entre eux figuraient déjà dans notre liste initiale. Ces noms, si nous ne les avons déjà recensés, auraient donc pu être aisément identifiés par la seule recherche des contextes du mot *culte*.

Quant au différentiel (356 noms propres absents de notre liste), il demande un examen plus attentif. Mis à part quelques locatifs (*culte d'Orient, d'Asie, d'Eleusis, de Karnak, de Rome, de Jérusalem, de Pessinonte, de Stonehenge...*), les noms propres désignent soit des dieux, soit des humains divinisés. On peut distinguer, plus précisément :

- des appellations que nous avons omises délibérément (*Yahvé, Jésus, Allah*, mais aussi *Eros* ou *Bel*), ou par inadvertance (une erreur technique avait fait disparaître le nom de *Bouddha* : l'interrogation du Web nous l'a rappelé à bon escient !);
- des noms de divinités rares, que nous n'avions pas trouvées dans notre documentation : par ex. *Bormanicus, Cernunnos, Dana, Ogma* (Celts), *Svantovit* (Slaves), *Nanâ* (Asie centrale), *Muruga* (Inde), *Tlatoc* (Aztlèques), *Fa* (Yorubas), etc. Ces identifications sont d'autant plus intéressantes qu'elles permettent de relancer la recherche et de rebondir, grâce aux liens constitutifs du Web, sur d'autres découvertes ;
- des désignations d'idoles, de « faux dieux » (*le culte de Mammon*) ;
- des variantes, phonétiques ou graphiques, de noms déjà connus (nous connaissons *Bastet* et *Asklépios*, mais non *Bâstet* et *Asklepios*). Il faut convenir toutefois qu'un certain nombre de graphies sont manifestement fautives (**Appollon*), même si la frontière entre fautes et variantes n'est pas toujours évidente ;
- des noms de « sous-divinités » (héros divinisés, personnages légendaires : *Achille, Cadmos, Hercule, Orphée*, mais aussi *Rama* ou *Bhîmasena*) ;
- des noms de personnages historiques qui ont fait l'objet d'un culte, à un titre ou à un autre : *Thoutmôsis, Néron, Robespierre, Napoléon, Atatürk, Mao* (ils partagent en partie les propriétés sémantiques et syntaxiques des noms de divinités).

Reste une dernière catégorie que nous n'attendions pas, et qui est propre à Internet : il s'agit de divinités « ludiques » – issues des jeux de rôles et des jeux vidéos –, dont la nomenclature est relativement bien représentée sur le Web : des noms comme *Bahamut, Cthulhu, Folo, Kostchtchie, Mask, Ogrimar, Sélène* (à côté de la classique *Séléné*), *Syl, Umberlee* semblent bien appartenir, sauf erreur de notre part, à ce monde occulte et mystérieux peuplé de monstres et de dragons, quand il ne s'agit pas de fausses divinités égyptiennes. Du point de vue linguistique, ils fonctionnent exactement comme les « vrais » dieux.

Le résultat, on le voit, est donc intéressant, même s'il comporte des aspects inattendus. Il est vrai que nous n'avons guère, par ce moyen, enrichi notre liste de divinités « traditionnelles » : quelques dizaines tout au plus. Mais ce n'est pas négligeable, compte tenu de la documentation encyclopédique dont nous étions partis. Encore s'agit-il là d'une expérience-limite. Si, pour une classe déjà décrite de manière approfondie, le Web permet de confirmer, de corriger, de compléter, à plus forte raison pourra-t-il contribuer à l'élaboration de nouvelles classes.

5. CONCLUSION

Les observations faites à propos des noms de divinités sont vraisemblablement transposables à l'ensemble des classes d'objets. Ici comme là les mêmes limites, mais aussi les mêmes richesses.

Les limites sont évidentes. Tout compte fait, la moisson récoltée sur le Web n'est pas très abondante, comparée à l'analyse que nous avons effectuée directement (et pour ainsi dire « manuellement ») dans notre première étude. Cette constatation est d'autant plus frappante que les moyens mis en œuvre sont relativement importants (des dizaines de milliers de pages, des centaines de milliers de concordances). Nous l'avons déjà dit, on ne peut tout attendre de l'automatisation : un amoncellement de données, si nombreuses soient-elles, ne remplacera jamais une véritable analyse, fondée sur des principes rigoureux et des textes soigneusement choisis.

Mais, à condition de s'articuler sur un travail linguistique, le Web peut contribuer efficacement à la description des classes : il constitue un réservoir d'exemples inépuisable ; une liste de noms étant donnée, il permet d'affiner l'analyse des contextes ; ou bien, réciproquement, partant d'une description contextuelle, il peut aider à l'identification de nouvelles unités lexicales. Pour une étude de la langue qui se veut centrée sur le lexique, la syntaxe et la sémantique, le Web constitue une source d'information irremplaçable.

BIBLIOGRAPHIE

- Daumas (F.), 1977 : *Les Dieux de l'Égypte*, 3^e édition, Paris, PUF, coll. Que sais-je ?
- Gross (G.), 1992 : « Forme d'un dictionnaire électronique », dans Clas (A.) et Safar (H.), édit., *L'environnement traductionnel, Journées scientifiques du réseau thématique Lexicologie, Terminologie, Traduction, Mons, 25-27 avril 1991*, Montréal, AUPELF-UREF, pp. 255-271.
- Le Pesant (D.) et Mathieu-Colas (M.), dir., 1998 : « Les classes d'objets », *Langages*, n° 131.
- Paumier (S.), 2004 : *Unitex 1.2*, <http://www-igm.univ-mlv.fr/~unitex>

LES PARTICULARITÉS LEXICALES DANS LES ROMANS, ENTRE DISCOURS ET FICTION

Omer Massoumou

département de langue et littérature françaises

Groupe de recherche en études sémantiques et interprétation (GRESI)

faculté des lettres et des sciences humaines, Université Marien Ngouabi

Brazzaville, République du Congo

***Résumé :** Dans les romans de notre corpus, l'usage des mots du terroir implique une prise en compte du contexte chez les auteurs. Au-delà des questions lexicographiques que soulève l'écriture des particularités lexicales, il est fait allusion aux problèmes lexicologiques, de définition sémantique acceptable. La lecture de ces textes littéraires suppose une prise en compte de certains critères pour valider le sens. La mise en exergue des processus, du parcours de l'émergence sémantique fait l'objet de ma réflexion.*

***Mots-clés :** particularités lexicales, français de référence, contexte littéraire, définition lexicale*

1. INTRODUCTION

En choisissant de parler des particularités lexicales dans les romans, nous voudrions nous placer tout de suite dans l'optique des études du français d'Afrique à partir des textes littéraires. Nous portons notre attention sur les africanismes qui, lorsqu'ils renvoient à un pays précis, prennent d'autres appellations. Les particularités lexicales sont reconnues comme telles en raison d'un écart vis-à-vis de la norme, celle du français de référence (FrR). Les xénismes sont aussi compris comme une forme de particularités lexicales. Ils renvoient à des formes différentes d'emprunts aux langues locales (africaines). Le concept de particularité lexicale concerne encore, comme le note Danièle Latin (1995 : 161) « la perception de faits complexes, de variété de langue ». Par ailleurs, il semble que les textes littéraires, les romans surtout, rendent compte d'une double attitude caractérisée par le besoin de restitution d'une certaine réalité et la nécessité de se situer dans un univers fictif. Ils mettent en exergue la relation consciente ou inconsciente qu'un auteur peut avoir avec les langues de son univers social. Il sera donc question de déterminer ici ce que représentent les particularismes dans un roman. C'est dans la lexicologie différentielle qu'il faut rechercher les outils d'analyse des lexies du français d'Afrique. Et comme le sujet abordé n'est pas réductible à cet aspect, nous allons prendre en compte les aspects de la linguistique énonciative qui renvoient au contexte d'énonciation. La différence entre discours et fiction telle qu'établie par certains chercheurs comme Gérard Genette, Tzvetan Todorov, Antoine Compagnon... servira à nourrir la réflexion sur le côté culturel ou stylistique de l'usage des particularités lexicales dans un récit fictif.

Pourquoi dans le cadre d'un récit romanesque un auteur choisit-il d'employer des particularités lexicales ? S'agit-il d'une expression naturelle, sans recherche esthétique ? Le français de référence n'offre-t-il pas de termes pour exprimer des faits propres à un espace culturel africain ?

Les questions relatives à ce type de réflexion relèvent de la part de réalité ou de la part de fiction que des mots précis donnent au récit. L'emploi des lexies particulièrement connotées représente en effet un enjeu dans les aspects lexicographiques, lexicosémantiques, stylistiques et culturels. Les variations linguistiques issues des pratiques linguistiques propres à des espaces plurilingues connaissent une existence certaine ou nouvelle dans le texte littéraire. Le contexte littéraire établit une vraisemblance historique qui peut empêcher de considérer les choses comme vraies mais qui ne permet pas non plus de les disqualifier. Une conscience linguistique des différences ou nuances entre les langues aboutit généralement à des traductions riches de sens.

De manière essentielle, nous portons notre intérêt aux formes et sens de certaines lexies dans *Sur la Braise* et *Le Mort vivant* d'Henri Djombo et dans *Histoire d'Awu* de Justine Mintsa. Comme ces auteurs ont écrit sur des faits relatifs au Congo (Brazzaville) et au Gabon, nous mettrons à profit les définitions apportées par les inventaires lexicaux de ces deux pays. Nous allons montrer, aussi bien au niveau lexicographique qu'au niveau lexico-sémantique, que les auteurs cités favorisent une « déconstruction lexématique et sémantique ». Pour démontrer que l'usage des particularités lexicales dans les textes littéraires répond davantage à un choix stylistique qu'à des contraintes culturelles, des exemples tirés de *Retour à Brazzaville* de Jean de Puytorac permettront de le confirmer.

2. CORPUS D'ÉTUDE

Mon *corpus* d'étude est constitué d'un ensemble de quatre romans. Les données sur la langue française ne seront pas étudiées de façon quantitative parce que le *corpus* n'est pas informatisé.

2.1. Des particularités lexicales renvoyant à des réalités du Congo

Nous présentons ici quelques citations contenant des congolismes. Pour les romans d'Henri Djombo, les extraits sont suivis des initiales du titre du roman et du numéro de la page où la citation apparaît.

Sur la braise décrit le parcours socioprofessionnel de Joseph Niamo qui mène une lutte sociale pour tenter d'apporter des solutions vitales dans une ville, Bocaville, et un pays, Boniko, où tout semble conduire vers le chaos, où l'arbitraire et l'irrationnel dominent :

- « La tutelle ne perçut dans la démarche du directeur général que le déni de son autorité irréfragable ; elle n'y vit qu'une manœuvre de déstabilisation : un dégé tentait de renverser un ministre ! » (p. 8) ;
- « Dans les langues du pays, on appelait ces moyens rapides des *golo-golo*. Pleins comme un œuf, ils ressemblaient à des cercueils mobiles. On reconnaissait les *golo-*

golo à leur couleur bleu blanc, aux slogans publicitaires de toutes sortes et aux inscriptions qu'ils portaient sur les côtés, à l'arrière et à l'avant » (pp. 28-29) ;

- « Alors, mon cher Jeff, il n'y a pas moyen d'être en règle avec les infractions que les *pourriciers* nous trouvent à tout bout de champ » (pp. 31) ;
- « Je vous demande aussi de travailler l'esprit du ministre afin qu'il me nomme à la place de cet homme » (pp. 119) ;
- « La réalité du pays sur ce plan est vraiment triste ! on appelle tous ces gens des *bi-boubous*, c'est-à-dire des idiots ou des naïfs du système. [...] Ces malheureux qu'on appelle aussi des *oranges-sucées-jetées* se demandent si, un seul instant de leur vie, ils ont été utiles à la société » (pp. 147).

Le Mort vivant porte sur l'arbitraire de la violence politique. Le héros qui sort des geôles de la dictature apporte à travers une longue lettre un témoignage sur les faits vécus. Les extraits ci-après présentent quelques particularismes lexicaux révélateurs :

- « Il n'y avait rien à craindre là-bas où ne planait aucune ombre policière. D'ailleurs, même s'il en avait plané, qu'aurait fait en pareille circonstance un *mange-mille*, ainsi qu'on appelle chez nous le policier, habitué à extorquer mille *séfás* aux usagers de la route [...] » (p. 23) ;
- « Il parla de la femme qui faisait de cet homme la risée de tous, tant elle affectait de forniquer avec des jeunes gens qu'elle appelait petits poussins » (Imv, 27) ;
- « Dans son état, il écoutait et voyait à peine les *nganga* et autres spécialistes appelés pour conjurer son malheur, [...] » (p. 174) ;

En dehors ces deux romans d'Henri Djombo, nous avons eu recours à *Une Vie au Congo, tome II : Retour à Brazzaville*, de Jean de Puytorac. Ce roman présente les aventures du narrateur dans le Congo de la période coloniale, le Moyen-Congo. L'administration coloniale, les actions des entreprises coloniales, les conditions de vie de la population autochtone... sont décrites dans une langue marquée par la couleur locale. L'auteur français Jean de Puytorac utilise des congolismes qui permettent d'avoir une certaine idée du français pendant la période coloniale au Moyen-Congo. Dans le cadre du récit les lexies sont contextualisées c'est-à-dire définies par l'auteur et elles à leur tour contextualisent le récit, le marquent par une couleur locale :

- « Le traitant est parti. La Compagnie belge est partie. Nous regrettons beaucoup *Mwana Moké* (petit enfant), M. Audier. Il est resté très longtemps avec nous, ici. Un jour, il a tout vendu à la Compagnie belge, et il est parti à M'Poutou. » (p. 25) ;
- « Beaucoup de perles, beaucoup de *pili* (toile de coton bleu), d'*américani*, des *bis-sengos* (houes), des machettes » (p. 25) ;
- « Il était onze heures, et le soleil donnait. Pas un pousse, pas une voiture devant l'institut Pasteur. Je pris "mon pied la route". Je passais devant une boutique lorsque je fus pris d'un éblouissement, et je m'évanouis sur place » (p. 105) ;

- « Alors, je lui enlevai son petit pagne, découvrant les multiples rangs de grosses perles, les *dzikitas*, qui lui ceignaient les reins et qui cliquetaient à chacun de ses mouvements » (p. 151) ;
- « Je n'ai pas eu la chance pour la médaille militaire. Elle au moins, te donne un petit peu de l'argent. Je touche la retraite du combattant... pas beaucoup de patas » (p. 254).

2.2. Des particularités lexicales renvoyant à des réalités du Gabon

À côté de ces différents textes, nous avons encore choisi de lire *Histoire d'Awu* de Justine Mintsá. Après avoir vécu la honte de la stérilité, Obame Afane fonde sa vie avec Awudabiran, une vie marquée par les petites et précieuses joies du bonheur familial et par des drames qui donnent une tonalité tragique au roman. Ce roman contient un nombre considérable de particularismes parmi lesquels :

- « Komandê, l'ancien commissaire de police, espérant toujours que sa pension lui parviendrait, tôt ou tard » ;
- « Or, je te le jure, après six ans de mariage, cette pauvre femme, personne ne la considérait plus comme belle-fille ou belle-sœur, M'bôm ou M'mieñ » (p. 40) ;
- « [...] je suis ta chose, comme tu me l'as rappelé, d'accord, mais je mange le foie de mon père si ta tête et la mienne se posent sur le même oreiller » (p. 105) ;
- « Quant à son esprit, elle (Awu) l'avait scellé en lieu inviolable » (p. 110) ;
- « Elle revoyait, avec une clarté que le temps ne parviendrait jamais à opacifier, un homme et une femme, manifestement heureux et pleins d'espoir, assis côte à côte dans un car de brousse bondé » (p. 110).

Ce *corpus* est nécessairement non exhaustif. Il offre néanmoins une possibilité de lecture de la nature de certains items retenus.

3. ANALYSE MORPHOLOGIQUE ET SÉMANTIQUE

L'analyse des particularités lexicales va nous permettre de définir leur aspect morphologique et sémantique. Le tableau récapitulatif ci-après facilite l'appréciation des différents termes du *corpus*. Les lexies se présentent sous des natures particulières : certaines lexies attestées dans le français local et d'autres sont forgées. Dans les quatre romans retenus pour cette étude, les emplois propres au français d'Afrique obéissent à cette double répartition. Le lecteur, quand il connaît l'univers social auquel renvoie le roman, est susceptible de faire une meilleure lecture. Quand il l'ignore, le processus d'interprétation sémantique devient moins aisé et plus approximatif. Une démarche de construction sémantique est alors entreprise et son aboutissement peut alors être plus ou moins heureux en fonction de plusieurs paramètres.

Type de particularités Romans	Particularités lexicales attestées dans la société	Particularités lexicales n'ayant de sens que dans le contexte textuel du roman
<i>Sur la braise</i>	<i>dégé</i> <i>travailler l'esprit</i>	<i>golo-golo</i> <i>pourriciers</i> <i>biboubous</i> <i>oranges-sucées-jetées</i>
<i>Le Mort vivant</i>	<i>mange-mille</i> <i>nganga</i> <i>petits poussins</i>	séfas
<i>Retour à Brazzaville</i>	<i>mwana moké</i> <i>m'poutou</i> <i>pili</i> <i>américani</i> <i>pousse</i> <i>dzikitas</i> <i>patas</i>	<i>prendre son pied la route</i> <i>bissengo</i>
<i>Histoire d'Awu</i>	<i>Komandé</i> <i>M'bom ; M'mien</i> <i>car de brousse</i>	manger le foie de son père sceller son esprit

3.1. Les particularités lexicales attestées dans la société

La connaissance du contexte ou de l'univers sociolinguistique de l'auteur permet de distinguer les lexies ayant un usage attesté. Certains vocables connaissent un usage dans les sociétés congolaise et gabonaise, d'autres non. Les écrivains ont tendance à recourir à une terminologie spécifique pouvant leur permettre de reproduire dans le texte littéraire la couleur locale de l'espace décrit. Pour y parvenir, le meilleur moyen, semble-t-il, est de reprendre des mots ayant une existence attestée dans le français du milieu. Les termes de la première colonne correspondent ainsi à des emplois communs mais non reconnus en FrR. Et dans ces conditions, l'usage qu'en fait un romancier consolide l'existence du mot. Pour ces différentes lexies, l'intégration lexicographique dans le système français ne pose pas de problème. Il existe une sémantaxe singulière qui demande à être exploitée ou expliquée selon le contexte du discours fictif.

Pour *dégé*, bien que la phrase révèle un emploi pas assez explicite, le lecteur comprend le sens général du discours. L'auteur fait une juxtaposition contextuelle qui permet au lecteur de noter qu'un *dégé* est un directeur général¹. La définition sémantique n'est donc pas assez explicite. Le terme *dégé* renvoie à l'acronyme de directeur général. La graphie proposée par l'auteur constitue une réponse lexicographique puisque d'autres variations graphiques sont connues : DG, D.G. L'usage d'un tel vocable dans le roman inscrit le texte littéraire dans l'espace sociolinguistique congolais. C'est le même cas avec le mot *pousse*. L'auteur Jean de Puytorac ne donne pas assez d'explication : « pas un pousse, pas une voiture devant l'institut Pasteur ». Le lecteur peut se demander s'il s'agit d'une énumération de moyens de déplacement ou d'une explication lexicale du terme *pousse*. La lecture des inventaires lexicaux du français au Congo permet de préciser avec Sandrine Veron qu'un *pousse* (variante graphique de *pousse-pousse*) était compris dans le contexte de la période coloniale du Moyen-Congo comme « une voiture légère à deux roues, tirée ou poussée par un homme et servant à transporter des personnes ou, plus rarement des marchandises » (1999, 33). A. Queffélec et A. Niangouna (1990, 258) parlent de chariot servant au transport des marchandises. En tenant compte du récit de Jean de Puytorac, c'est le sens défini par Sandrine Veron qui est plus pertinent. Le terme *pousse* renvoie à un contexte historique précis et ne peut être compris que par rapport à cela. En effet, le *pousse* servait au transport des personnes et des marchandises mais actuellement, il ne sert plus qu'au transport des marchandises. La forme morphosyntaxique *travailler l'esprit* est comprise dans le sens de *marabout* (nuire à autrui ou le contraindre de façon magique à agir dans un sens souhaité). Le contexte sociolinguistique permet donc de mieux comprendre le récit fictif même par des lecteurs ne partageant pas l'espace culturel de l'auteur.

Pour le mot *mange-mille*, l'auteur Henri Djombo fait une glose intéressante. « Un *mange-mille* ainsi qu'on appelle chez nous le policier, habitué à extorquer mille séfas aux usages de la route ». L'écrivain fait un travail de lexicologue. Il apporte des précisions sémantiques intéressantes pour le lecteur étranger à l'espace de référence du roman. Ce terme *mange-mille* ne figure pas dans les inventaires lexicaux du français au Congo².

Il y a donc une certaine imprécision lexicale dans les emplois lexicaux avec des mots comme *dégé*, *pousse* et *car de brousse* ou *américani* que nous ne développons pas ici. Par contre pour d'autres lexies, les auteurs font un véritable travail de lexicologue. Ils proposent des définitions sémantiques. *Mange-mille*, *petits poussins*... sont sémantiquement expliqués.

À côté de ces mots qui s'intègrent aisément dans le système phonéticographique du français s'en trouvent d'autres dont la graphie rappelle leur origine. Ce sont les emprunts aux langues congolaises ou gabonaises. Des mots comme *nganga*, *mwana moké*, *m'poutou*, *pili*, *dzikitas*, *patas*, *m'bom*, *m'mien*, *komandé* expriment une intention des auteurs de reproduire des réalités immédiates avec des mots beaucoup plus expres-

¹ Jean-Alexis Mfoutou (2000, 62) présente la même définition.

² On pense aux travaux de A. Queffélec et A. Niangouna (1990), J.-A. Mfoutou (2000), O. Massoumou (2001).

sifs. Les auteurs deviennent des lexicographes dans la mesure où ils contribuent à la conversion des « vocables en lexème » (Mortureux 2004 : 13). Les formes graphiques présentées dans les récits ne sont pas toujours celles qu'on retrouve dans des inventaires lexicaux où les variantes sont systématiquement ignorées. Le contexte littéraire tend donc à développer une seule forme lexicale pour tout mot à travers le récit. L'accentuation du *e* final dans *mwana moké*, le *m* apostrophe et la diphtongue « ou » dans *m'poutou* (on aurait pu avoir *mputu*, *m'putu...*), etc. montrent que les écrivains participent à l'émergence d'un français local ou régional. Et pour ces emprunts, ils présentent des définitions lexicales mises entre parenthèses (c'est le cas avec *pili*, *dzikittas*, *mwana moké*)³ ou des explications directes avec les lexies comme *nganga*, *patas*.

3.2. Les lexies non attestées dans les sociétés congolaise et gabonaise

Les quelque neuf lexies socialement attestées retenues de ce *corpus* permettent une analyse prospective des particularités lexicales dans la mesure où elles posent le problème de la référence lexicale et de l'arbitraire du signe linguistique. Nous distinguons deux types de particularités lexicales, celles qui ont une morphologie française (*oranges-sucées-jetées*, *prendre son pied la route*, *manger le foie de son père*, *sceller son esprit*) et celles qui viennent des langues africaines (*golo-golo*, *biboubous*, *bissengo*). Des lexies composites ou hybrides comme *pourriciers* et *séfas* complètent la liste des mots non attestés dans le milieu social.

L'examen de la réalité n'aurait probablement pas permis le recensement des particularités lexicales comme *golo-golo*, *pourriciers*, *prendre son pied la route*, *biboubous* et *oranges-sucées-jetées*. Ces particularités apparaissent dans un cadre textuel précis qui est le contexte romanesque.

C'est le cas de *sceller son esprit* et dans une moindre mesure de *pourriciers* et *oranges-sucées-jetées*. *Sceller son esprit* est compris dans le sens de se protéger contre les mauvais esprits, les sorciers. Le terme peut être admis au-delà du cadre romanesque mais il n'est pas suffisamment lexicalisé. Par contre, les autres lexies *golo-golo* et *biboubous* renvoient à une création lexicale de l'auteur. Elles intègrent une dynamique du fonctionnement de la fiction romanesque. Le vocable *golo-golo* rappelle par sa composition le terme *foula-foula* qui dans les langues congolaises désigne le minibus de transport en commun. Dans un passé encore récent, il fallait être suffisamment fort (*ngolo* en kituba) pour y accéder au niveau des arrêts de bus en raison des bousculades. La création littéraire exploite ici, par un brouillage sophistiqué, les éléments de la réalité quotidienne. La trame du récit donne une définition du vocable *biboubous*. Cette pratique du narrateur correspond à une démarche lexicographique relative à la lemmatisation, à la définition du sens d'un mot. Mais du fait que l'usage de ce terme n'est pas suffisamment lexicalisé dans le français au Congo, il n'est pas possible de le considérer comme particularité lexicale attestée.

³ Le dictionnaire bilingue de René van Everbroeck *Lingála-Français, Français-Lingála* propose la graphie *mpili* pour l'étoffe bleu foncé ou noire portée en signe de deuil (p.143). Moké est orthographié par *moké* (p. 123).

Le vocable *pourriciers* est un mot-valise utilisé par certains locuteurs congolais. Au départ, il s'agit d'une mauvaise réalisation phonétique du mot *policier* par des locuteurs basilectaux. La syllabe « li » est prononcée [Ri] et le terme devient « poricier ». Pour mettre en exergue le comportement déviationniste, donc pourri du policier, des intellectuels ont produit le vocable « pourricier ». L'auteur de *Sur la braise* reprend un mot qui a une certaine existence dans le français du Congo. Le vocable *oranges-sucées-jetées* a dans le roman le sens de personnes naïves ayant servi de marchepied à quelqu'un. Il s'agit d'une extension sémantique du mot qui d'ordinaire renvoie à des femmes avec qui un homme fait l'amour juste pour satisfaire sa libido.

Dans le cadre d'un inventaire lexical, la prise en compte de ces lexies dépend de plusieurs critères. Mais leur présence au sein d'un *corpus* peut permettre de les identifier ; ce qui ne garantit pas leur sélection définitive. La question de la lemmatisation reste suspendue à celle d'un large *corpus* qui garantirait l'usage aussi bien au niveau de la forme que du sens. Tous ces éléments permettent de dire que l'univers de fiction ne porte pas que sur l'intrigue mais aussi sur l'usage de la langue ou des langues d'écriture. Il existe un contexte qui peut aider à comprendre l'œuvre. Si Edema (2004 : 230) subordonne l'explication de l'œuvre à une connaissance approfondie des langues du texte africain, il convient de considérer le récit dans un contexte large qui associe le discours, la réalité du monde auquel le récit renvoie et la fiction même. C'est dans ce sens que le fonds culturel participe à l'interprétation du fait littéraire, même si l'auteur opère initialement un choix stylistique.

4. EMPLOI LEXICAL ET CONTEXTE ROMANESQUE PERTINENT

Le contexte romanesque peut être perçu dans une double démarche, soit il est pris dans le sens de la référence à la réalité soit il est compris dans la perspective du récit fictif. Antoine Compagnon (1998 : 141-147) évoque le déni de la « faculté référentielle de la littérature » dans la théorie générale française. « Comme il a toujours régné dans cette discussion, [...] une confusion certaine entre la référence dans la langue et l'école réaliste en littérature, identifiée au roman bourgeois, il n'est pas possible d'ignorer le contexte historique dans lequel la thèse de l'arbitraire de la langue a été reçue » (Compagnon 1998 : 147). Il est plutôt question ici de reconnaître les implications de la réalité dans le fait littéraire. Le texte littéraire est vraisemblable. C'est pour cela qu'en sa qualité de fiction, il « s'écarte [...] du discours référentiel en ce que ses phrases ne renvoient pas à des référents "réels" ». (Ducrot, Schaeffer, 1995 : 372). Et le discours est davantage à considérer dans le sens de « situation de discours » définie comme « l'ensemble des circonstances au milieu desquelles a lieu une énonciation écrite ou orale » (Ducrot, Schaeffer, 1995 : 764). La catégorisation lexicale a permis d'identifier deux formes de lexies. Elle ne prend pas en compte les éléments de la société à laquelle renvoie le texte littéraire. Il est donc question de porter une attention particulière sur ce qu'on peut appeler le « contexte pertinent » : « le contexte pertinent pour l'étude littéraire d'un texte littéraire, ce n'est pas le contexte d'origine de ce texte, mais la société qui en fait un usage littéraire en la séparant de son contexte d'origine » (Compagnon 1998 : 49). La réflexion débouche ainsi sur les aspects de la théorie de la réception du fait littéraire. Pour comprendre les particularités lexicales dans un univers fictif, il n'est

pas forcément question de partager l'univers social évoqué. Des textes de la littérature francophone d'importance majeure comme *Texaco* (Patrick Chamoiseau), *Pélagie-la-charrette* (Antonine Maillet) sont fort marqués par l'emploi des particularismes lexicaux. Le lecteur, en sa qualité de « faiseur de sens », pose les sens pertinents par rapport au contexte. À partir des éléments textuels présentés par le narrateur, il y a définition d'orientations sémantiques pertinentes. Les lexies attestées dans un espace linguistique donné peuvent ainsi être utilisées par n'importe quel locuteur de la langue française. Et leur emploi dans une œuvre de fiction intègre une stratégie d'écriture dès lors que l'on se rend compte que l'enjeu du français en contexte africain participe davantage à la structuration de l'intrigue. C'est le cas avec *Allah n'est pas obligé* où le narrateur organise son récit en jouant à la fois sur les particularismes linguistiques et sur les plurilinguismes ambiants. Le roman *Une Vie au Congo, tome II : Retour à Brazzaville*, est intéressant à ce titre. L'auteur aurait pu nous proposer un texte exempt de congolismes puisqu'il est français. Sa compétence linguistique lui aurait permis de traduire les faits du contexte d'origine par des mots dont l'emploi resterait en harmonie avec le français de référence. Nous pensons à André Gide qui, dans son journal de voyage, *Voyage au Congo* (1926) ne recourt pas de façon significative aux mots du français local. Entre André Gide et Jean de Puytorac, nous constatons qu'il y a deux stratégies différentes d'écriture. L'un choisit d'écrire en français local, l'autre préfère garder les mots du FrR.

La conscience lexicologique exprimée par les auteurs des quatre romans de notre *corpus* procède par explication ou traduction des lexies non conformes au FrR. Tout lecteur peut arriver à comprendre le sens du texte des particularités lexicales grâce aux indications lexicologiques apportées par les auteurs. Mais dans certains cas, le texte peut échapper à la logique d'explication sémantique. Le lecteur considère alors des hypothèses qui peuvent ne pas correspondre à un sens pertinent. Dans *Une Vie au Congo, tome II : Retour à Brazzaville* de Jean de Puytorac, l'emploi des congolismes comme *bissengos* ou *pili* pose le problème de l'écriture et de la traduction des particularismes. Ces deux lexies ont une forme peu habituelle ; la graphie de *bissengos* est *bassengo* et *pili* tiré du lingala s'écrit *mpili*. Les déformations graphiques de ces mots peuvent être interprétées de deux manières. Soit il est question d'une volonté de l'auteur effet de sa fantaisie, soit l'auteur utilise une forme mineure peu répandue. En effet, *nsengo* fait son pluriel en *bassengo* (*s*) et non *bissengos*. Le terme *pili* doit avoir une consonne initiale nasale *mpili*. Ces écarts morphologiques rendent compte de l'enjeu du contexte. Tous les locuteurs d'une langue ne conçoivent pas les faits linguistiques de la même manière.

Henri Djombo propose des termes comme *séfas*, *pourriciers*... La lexie *séfas* apparaît dans un contexte où il est aisé de faire un rapprochement au CFA, au franc de la communauté financière d'Afrique. Le choix morphologique de l'auteur correspond davantage à une fantaisie, à une créativité ambiguë de l'auteur qu'à une pratique bien connue.

5. CONCLUSION

Nous venons de voir que le contexte littéraire se caractérise par deux types de particularités lexicales. D'une part, les écrivains emploient des lexies propres à un espace linguistique donné, le Congo ou le Gabon, d'autre part, ils produisent des lexies qui n'ont de sens que par rapport au contexte textuel. Ces différents particularismes sont sémantiquement différents des mots du français de référence. Cette réalité justifie les démarches d'explication ou de traduction lexicale qu'adoptent les écrivains. Dans les romans étudiés, nous avons noté que les lexies jouent une double fonction, celle d'exprimer des faits et celle de construire l'intrigue. L'emploi des particularités lexicales donne donc une valeur ajoutée au texte littéraire.

Une recherche portant sur toute la littérature congolaise ou sous-régionale, à partir d'une base de données informatisées, serait intéressante à mener pour pouvoir établir un vocabulaire de la littérature et enrichir les inventaires lexicaux du français dans les différents pays de l'Afrique centrale. La lecture des textes littéraires permet de trouver des illustrations d'articles des inventaires lexicaux. Et dans la perspective de la construction de la base de données panfrancophone au niveau de l'Afrique centrale, la problématique des particularités lexicales dans le discours de fiction peut être poursuivie dans la mesure où elle est en mesure de favoriser des réponses originales au sujet des variétés de français. Le français écrit est davantage pratiqué par les lettrés.

BIBLIOGRAPHIE

- Boucher (K.) et Lafage (S.), 2000 : « Le lexique français du Gabon, entre tradition et modernité », *Revue Le Français en Afrique* n° 14, Nice, ILF-CNRS, 415 pp.
- Chamoiseau (P.), 1992 : *Texaco*, Paris, Gallimard, 504 pp.
- Compagnon (A.), 1998 : *Le Démon de la théorie*, Paris, Éditions du Seuil, « Points », 341 pp.
- De Puytorac (J.), 1995 : *Une vie au Congo, tome II : Retour à Brazzaville*, Cadeilan, Éditions Zulma, 317 pp.
- Djombo (H.), 2000 : *Sur la braise*, Brazzaville, les Éditions Hémar, deuxième édition, 201 pp.
- Ducrot (O.) et Schaeffer (J.-M.), 1995 : *Nouveau dictionnaire encyclopédique des sciences du langage*, Paris, Éditions du Seuil, 821 pp.
- Edema (A. B.), 2004 : « Les xénismes dans les romans africains : entre citations, traduction et créativité lexicale », dans *Le français en Afrique, revue du réseau des observatoires du français contemporain en Afrique*, n° 19, Nice, Institut de linguistique française – CNRS, pp. 227-243.
- Everbroeck (R. V.), 1985 : *lingála-français, français-lingála*, Kinshasa, Éditions l'Épiphanie, 258 pp.
- Kourouma (A.), 2000 : *Allah n'est pas obligé*, Paris, Le Seuil.
- Latin (D.), 1995 : « Particularités lexicales et variétés du français : l'enjeu de l'instrumentalisation lexicographique. À propos des inventaires du Sud et du français en Afrique noire », dans Francard (D.) et Latin (D.), *Le Régionalisme lexical*, Louvain-la-Neuve, Duculot, pp. 159-170.
- Maillet (A.), 1990 : *Pélagie-la-Charrette*, Québec, Leméac éditeur, 334 pp.
- Massoumou (O.), 2001 : « Pour une typologie des néologies du français au Congo-Brazzaville », dans *Le français en Afrique* n° 15, pp. 133-168.
- Mfoutou (J.-A.), 2000 : *Le Français au Congo-Brazzaville*, Maromme, Éditions Espaces culturels, 302 pp.
- Mortureux (M.-F.), 2004 : *La Lexicologie entre langue et discours*, Paris, Armand Colin, 192 pp.

- Ngamountsika (E.), 2004 : « De l'écriture romanesque à la créativité lexicale. Lecture de *Sur la braise* et *Le Mort vivant* d'Henri Djombo », dans *Cahiers du GRESI*, Brazzaville, revue du département de Langue et littérature françaises, pp. 53-67.
- Queffélec (A.) et Niangouna (A.), 1990 : *Le français au Congo*, Aix-en-Provence, PUP, 333 pp.
- Veron (S.), 1999 : « Le français au Moyen Congo à l'époque coloniale (1920-1940) : inventaire lexical », dans *Le Français en Afrique* n° 13, InaLF-CNRS, Paris, Didier Érudition, pp. 6-38.

MORPHOLOGIE DE SPÉCIALITÉ : REGARD(S) SUR LE(S) CONTEXTE(S)

Raquel Silva

centre de linguistique
Universidade Nova de Lisboa
Lisbonne, Portugal
raq.silva@fcsh.unl.pt

Résumé : *Tout ou presque en linguistique est une question de perspective. L'angle sous lequel nous avons choisi d'exposer nos propos repose sur une méthodologie d'orientation textuelle où le corpus spécialisé assume un rôle prépondérant dans l'organisation des connaissances, et ceci en fonction des progrès de la terminologie computationnelle, en matière de développement de logiciels pour l'extraction de termes. L'abordage linguistique à la terminologie textuelle pose actuellement de nouveaux défis d'ordre théorique et méthodologique, nous nous attardons sur ce concept récent en terminologie pour, ensuite, rendre compte de l'observation de phénomènes morphologiques particuliers, rencontrés en terminologie médicale. Nous cherchons ainsi à contrarier l'idée que ce qui différencie la langue de spécialité de la langue générale est à peine d'ordre lexical, posant ainsi l'hypothèse d'une morphologie de spécialité. De par leurs spécificités de composition morphologique, les unités terminologiques en analyse, outre leur fonction dénominative, constituent en soi des microcontextes riches et particulièrement propices à l'établissement de paradigmes distributionnels de formants, et ceci en contexte de langue de spécialité.*

Mots-clés : *terminologie textuelle, morphologie de spécialité, microcontexte morphologique, formant, terme, dénomination*

1. INTRODUCTION

La terminologie est entrée dans le XXI^e siècle en posant ses jalons de discipline scientifique. Le débat étant centré, du côté des linguistes terminologues, sur son appartenance ou non à la linguistique. L'heure est donc aux bilans théoriques, aux parcours historiques, à l'état des lieux, le moment étant venu de retrouver ses marques par rapport à une discipline aux multiples facettes et à la croisée de divers champs d'application. Les thèmes des derniers colloques¹ et les titres des récentes publications² sont très révélateurs sur cette nouvelle phase d'introspection que traverse la terminologie.

¹ *Workshop sur la théorie de la terminologie* (Prague, 2003), LSP européen (Surrey, 2003), *Terminologie, l'état des théories* (Lisbonne 2003) ; *La terminologie, nouvelle discipline scientifique* (Paris 2003), etc.

² *Terminologie : nouvelles orientations* (RFLA 1998) ; *La Terminologie, discipline scientifique* (SFT 2004) ; *Les Fondements théoriques de la terminologie* (Cahiers du Ciel 2004) ; *La Terminologie : nature et enjeux* (Langages 2005), *Terminologie, l'état des théories* (Termip, CLUNL 2005), etc.

Diverses façons d'aborder la discipline sont aujourd'hui possibles en fonction des objectifs prétendus, qu'ils soient de nature exclusivement théorique ou non, orientés vers la création de ressources terminologiques, le traitement automatique des langues, les technologies de l'information, la modélisation des connaissances ou encore la planification linguistique.

Autant de champs d'application font de la terminologie une discipline aux multiples enjeux comme l'affirme Depecker (2005) ; enjeux industriels, économiques, scientifiques, culturels et politiques. Une chose est certaine, l'enthousiasme pour la discipline est en crescendo et les débats scientifiques dont elle est le centre des attentions lui confèrent, sans la moindre équivoque, un statut particulier au sein des études sur et à partir de la langue et/ou du discours.

C'est donc dans la lignée des études linguistiques portant sur les unités terminologiques que nous inscrivons notre démarche. Celle-ci repose sur une méthodologie d'orientation textuelle où le *corpus* spécialisé assume un rôle prépondérant dans l'organisation des connaissances.

Dans ce sens nous aborderons en premier lieu le concept de « terminologie textuelle », afin d'insister sur l'importance de la préparation du *corpus* dans l'abordage textuel à la terminologie. Puis, nous présenterons quelques éléments de réflexion portant sur l'étude de spécificités morphologiques rencontrées en langue de spécialité, ayant pour base un *corpus* de médecine, et ceci dans le but d'apporter une contribution supplémentaire à l'approche linguistique de la terminologie : démontrer l'existence d'une « morphologie de spécialité ».

2. TERMINOLOGIE TEXTUELLE

La révolution communicationnelle qui s'est opérée au sein de la société de l'information, à la fin du XX^e, a clairement fait émerger de nouvelles nécessités sur le terrain de la terminologie. De nouvelles orientations ont vu consolider leurs théories et leurs méthodologies, en fonction de besoins terminologiques très précis. Celles-ci se développent à des fins strictement scientifiques ou bien auprès de secteurs d'activités très diversifiés, entreprises ou institutions, de plus en plus conscientes des dividendes possibles d'une terminologie facilement accessible, de qualité et bien organisée.

Le champ d'application est aujourd'hui d'autant plus vaste que les besoins d'échanges multilingues sont un phénomène de mondialisation incontournable, la terminologie n'y échappant pas ! C'est dans ce récent contexte socioéconomique que la terminologie trouve de nouvelles pratiques, la terminologie textuelle en est le plus récent exemple.

Contaminée positivement par l'influence des théories sur les linguistiques de *corpus*, la terminologie a redécouvert le texte. D'après Condamines (2005 : 36), « Une des conséquences de cette évolution est que la terminologie en tant que discipline scientifique s'est rapprochée de la linguistique ». En effet, les terminologues habitués à explorer les textes pour en repérer les concepts à travers leurs désignations, portent maintenant aussi leurs regards sur les frontières du terme pour mieux percevoir la nature des relations linguistiques et conceptuelles qui caractérisent le texte de spécialité comme un tout, dans un domaine précis.

Vu sous cet angle, la linguistique elle-même finit également par bénéficier de ce rapprochement. En effet, la perspective apportée par la terminologie textuelle ouvre de nouveaux chemins à l'analyse linguistique, surtout si l'on considère comme Condamines (2005 : 42-43) que « la terminologie textuelle se rapproche de l'analyse de discours qui considère que toute analyse de textes est un construit et est soumis à une interprétation ».

Face à l'importance croissante et inévitable des *corpus* spécialisés comme source de dépouillement terminologique, des avancées théoriques récentes vont plus loin et centrent leur attention non plus seulement sur les termes, les frontières du terme, mais aussi sur les limites du texte, et suggèrent l'établissement de typologies textuelles spécialisées. Ce point de vue est justifié comme cela par Costa (2005) : « Tous les *corpus* ne servent pas le même but, ce qui nous oblige à réfléchir sur les critères de composition de *corpus* et en conséquence à établir des critères de sélections de typologies de types et de genres de textes et/ou de discours. Car, ce n'est qu'en fonction de l'établissement de ces critères qu'il est possible de garantir l'adéquation des textes aux objectifs préétablis qui sont évidemment le garant de tout travail de recherche ».

Les textes spécialisés passent ainsi de simples réceptacles de termes à des ensembles organisés et structurés, voir à ce titre la thèse de doctorat de Rute Costa (2001) ainsi que ses réflexions, menées au Centre de linguistique de l'Universidade Nova de Lisboa-CLUNL³ dans ce domaine. Les arguments fondamentaux apportés ont pour base l'insuffisance des critères de classification préétablis pour la constitution de *corpus* terminologiques, puisque, selon Costa (2005) : « Toute approche terminologique qui se préoccupe de la qualité de sa matière première, dans ce cas précis, la production textuelle et/ou discursive, a plus de garanties d'obtenir des résultats efficaces qui seront fort utiles à la manutention de la qualité linguistique nécessaire à la construction de toute ressource linguistique et essentielle à toute communication ».

La simple et traditionnelle classification des textes en domaines reste certes fondamentale mais n'est plus suffisante, d'autres critères d'ordre sociologique et sociolinguistique viennent s'y ajouter, comme l'identification des communautés scientifiques productrices et réceptrices de ces textes, la diversité des situations de communication et l'encadrement spatio-temporel de leur diffusion. Ce sont ces paramètres qui permettent l'interprétation des données terminologiques contenues dans les textes spécialisés.

À ces critères viennent naturellement s'associer des critères d'analyse linguistique basés sur l'étude approfondie des unités terminologiques en contexte, on parle aujourd'hui en terminologie textuelle d'une sémantique de *corpus* propre à l'établissement d'arborescences terminologiques. Cette perspective d'analyse est en effet entrée dans la pratique terminologique par la voie des systèmes experts de l'ingénierie des connaissances, à travers la manipulation de grands *corpus* spécialisés et ayant comme finalité l'organisation des connaissances contenues dans les textes.

Il est aujourd'hui démontré que la voie du traitement automatisé de la langue, indépendamment de la spécialisation linguistique, qu'elle soit lexicologique, lexicographique, syn-

³ Unité de recherche LI2 « Lexicologie, lexicographie et terminologie ».

taxique, sémantique ou morphologique, a forcément des répercussions sur le renouvellement des théories sous-jacentes à ces disciplines et la terminologie en est actuellement un bon exemple.

Depuis quelques années déjà, nous travaillons au CLUNL sur les *corpus* de spécialité, et l'un des principaux domaines de recherche en terminologie porte sur la langue de la médecine⁴. Les sous-*corpus* actuellement compilés dans différentes branches comme l'anatomie, la sénologie, la cardiologie ou l'imagerie médicale ont tous fait l'objet de thèses en linguistique/terminologie, déjà conclues ou en cours.

La médecine est un domaine de recherche fondamental du point de vue terminologique puisque, au Portugal, la majeure partie des textes scientifiques médicaux produits, le sont, originalement, en langue étrangère et principalement en anglais. Cette situation oblige donc les étudiants à accéder aux textes, et par conséquent aux concepts et aux termes, en langue étrangère ou bien à partir d'une traduction, ce qui parfois se révèle être bien plus dangereux ! Et ceci précisément, au moment où l'on parle de plus en plus du concept de « perte des domaines » comme étant l'incapacité à s'exprimer dans sa langue à tous les niveaux de spécialité, justement par manque de terminologie.

Les méthodologies développées pour la terminologie textuelle, recourant aux critères de classification, organisation et structuration de *corpus* dont nous avons parlé antérieurement peuvent être un excellent outil, non seulement pour la préservation des textes eux-mêmes mais surtout de leur contenu. Ils peuvent ensuite être utilisés à plus d'une finalité, que ce soit à la construction de ressources terminologiques dans le but de stabiliser l'usage de certains termes ou l'automatisation des langues de spécialités en vue de leur diffusion à plus large échelle et en langue originale.

3. TERMINOLOGIE ET MORPHOLOGIE

Le *corpus* que nous exploitons actuellement dans le cadre de la thèse est un *corpus* monolingue dans le domaine de l'imagerie médicale. Ce *corpus* réunit une compilation de textes scientifiques servant de bibliographie de base à cette discipline dans les cursus universitaires de médecine, au Portugal. Ces textes hautement spécialisés sont rédigés par des médecins spécialistes qui, pour la plupart, sont aussi professeurs de leurs disciplines et dont la préoccupation est également de transmettre à leurs étudiants la terminologie portugaise du domaine étudié.

Nous sommes d'avis que plus les textes spécialisés sont objets de fines analyses linguistiques, plus ils sont valorisés par la propre communauté de spécialistes qui les produit. N'ayant très souvent pas conscience des implications futures, au moment du choix des termes pour s'exprimer, ils se retrouvent sensibilisés par la préoccupation des terminologues lorsqu'ils comprennent que leurs « langages » sont entre nos mains et sont, généralement, de bons collaborateurs quand il s'agit de valider leurs terminologies.

⁴ Projet sous la direction scientifique de Teresa Lino.

L'aspect scientifique qui a dès le départ motivé notre intérêt d'étude, bien avant de recueillir le *corpus* de l'imagerie, était la description de phénomènes néonymiques dans les textes de spécialité et principalement le passage entre le réel, moment de la création de concept, et son expression en discours. Il nous fallait pour cela un *corpus* qui reflète cette dynamique dénominative, un domaine scientifique hautement technologique s'imposait, car généralement, les progrès rapides des deux branches et leur fusion, en matière de recherche, ont des conséquences terminologiques immédiates au niveau de l'activité néonymique des dénominations.

Nous cherchions également à contrarier l'idée, longtemps cristallisée en terminologie, que ce qui différencie la langue de spécialité de la langue générale est à peine d'ordre lexical et que du reste, elles partagent les mêmes règles de fonctionnement syntaxique et de formation morphologique. De fait, « en plus des études portant sur les catégories propices à la formation de termes (nom, verbes, adjectifs et adverbes), s'amorcent aussi les recherches sur la structure interne des unités terminologiques, puis sur les combinaisons larges ou plus restreintes au sein des langues de spécialité (Lsp). De surcroît, le niveau d'analyse morpho-syntaxique – parfois réduit au rôle de microcontexte du terme et jusqu'alors dans l'ombre, en faveur des niveaux sémantique, morphologique, pragmatique et conceptuel – se retrouve au premier plan. » (Silva, Costa et Ferreira 2004).

Dans cette perspective et donnant suite aux travaux déjà en cours, en terminologie textuelle sur les « collocations terminologiques » (Silva, Costa et Ferreira 2004) et en terminologie computationnelle sur les « collocations terminologiques verbales » (Costa et Silva, 2004), nous avons orienté notre recherche sur la relation entre terminologie et morphologie dans un *corpus* comme celui de l'imagerie où circulent des unités terminologiques très particulières. Il s'agit de termes dont la construction morphologique se caractérise par une juxtaposition d'éléments savants, pouvant aller jusqu'au nombre de cinq comme par exemple dans *colangiopancreatografia* ou *hemangiomiélolipoma*.

L'étude d'un procédé de composition morphologique tel que celui-ci en terminologie, c'est-à-dire la formation de monoterme par concaténation de formants morphosémantiques, nous paraît un indice révélateur de phénomènes linguistiques propres aux langues de spécialités. La spécialisation morphologique de la terminologie a permis de soulever quelques problématiques intéressantes, notamment sur la non-applicabilité de l'analyse morphologique traditionnelle, puisque les notions de radical, base, préfixation, suffixation et composition sont à reconsidérer en fonction de facteurs de créativité terminologique pour répondre à des besoins dénominatifs. Par ailleurs, de par leurs spécificités de composition morphologique, les unités terminologiques en analyse, outre leur fonction dénominative, constituent en soi des microcontextes riches et particulièrement propices à l'établissement de paradigmes distributionnels de formants, et ceci en contexte de langue de spécialité.

L'un des aspects les plus intéressants est précisément le mécanisme linguistique qui s'opère pour rendre compte de ses besoins dénominatifs. L'activité dénominative déclenchée par un facteur d'ordre extralinguistique, puisqu'il s'agit d'un nouveau concept, se traduit par une dynamique de la langue, une capacité à attribuer un nom à une chose. La répercussion linguistique au niveau de la construction des termes assume divers mécanismes. L'insertion de formants entre la tête et la terminaison du terme est l'un des procédés morphologiques

les plus productifs dans notre *corpus*. Prenons par exemple le terme *angiografia* par rapport à *angiocardipneumografia* : l'insertion de nouveaux éléments provoque l'expansion morphosémantique de la désignation du terme, créant un nouveau concept, et reflétant ainsi les progrès de la science et l'avancée des technologies en imagerie.

D'autres mécanismes comme la suppression de formants à l'intérieur des termes montre que généralement la syncope d'un élément savant laisse des traces au niveau du concept et ceci n'est perceptible qu'à travers la définition du terme. Par ailleurs, « Ce procédé n'est pas très fréquent mais il s'assimile en quelque sorte à un autre phénomène bien connu, celui de l'économie de la langue. Dans ce cas, la suppression du formant est justifiée par son absorption sémantique et conceptuelle au niveau des autres formants qui composent le terme. Ce phénomène a d'autres implications terminologiques, notamment celle de créer la co-occurrence de termes synonymes » (Silva 2005). Quant à la mobilité des formants à l'intérieur des termes, une autre des caractéristiques observées, par exemple pour le formant Chole qui apparaît tantôt en position initiale, *cholangiografia*, tantôt en position centrale du terme, *angiocholégafia*, ce phénomène multiplie malgré lui les synonymes et les variantes terminologiques et ceci dans un domaine médical où la précision du langage, fruit de l'observation méticuleuse de l'image, se révèle être une règle d'or parfois difficile à accomplir.

4. CONCLUSION

Les concepts de « dynamique dénomminative » et de « productivité morphologique » (Silva 2005) sous-jacents à l'étude des termes de l'imagerie médicale sont des phénomènes linguistiques qui permettent de constater une activité néonymique très forte au sein d'une terminologie en constante évolution scientifique et technologique. Nous sommes d'avis que l'étude en cours, à partir de *corpus*, est venue enrichir les recherches sur les spécificités de la langue de la médecine, aussi bien du point de vue terminologique que linguistique, dans la mesure où de nouvelles pistes d'analyse morphologique vont contribuer à la description de phénomènes entrant dans la construction de certains termes hautement spécialisés, posant ainsi l'hypothèse de l'existence d'une morphologie de spécialité.

BIBLIOGRAPHIE

- Condamines (A.), 2005 : « linguistique de *corpus* et Terminologie » dans *Langages 157, La Terminologie : nature et enjeux*, Paris, Larousse, pp. 36-47.
- Depecker (L.), 2005 : « Contribution de la terminologie à la linguistique » dans *Langages 157, La terminologie : nature et enjeux*, Paris, Larousse, pp. 6-12.
- Costa (R.), 2001 : *Pressupostos teóricos e metodológicos para a extracção automática de unidades terminológicas multiléxicas*, thèse de doctorat en linguistique, Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa, Lisboa.
- Costa (R.), Silva (R.), 2004 : « The Verb in the Terminological Collocations. A Contribution to the Development of a Morphological Analyser MorphoComp » dans Lino (T.), Xavier (F.), Ferreira (F.), Costa (R.) et Silva (R.), édit., *IV LREC Proceedings*, vol. IV, Lisboa, ELRA, pp. 1531-1534.

- Costa (R.), 2005 : « Terminology, *corpora* linguistics and ontologys » dans TKE 2005 – 7th International conference on Terminology and Knowledge Engineering. Terminology & Content Development, Copenhagen Business School, Denmark (sous presse).
- Silva (R.), Costa (R.), Ferreira (F.), 2004 : « Entre langue générale et langue de spécialité : une question de collocations » dans ELA, *Vocabulaires de spécialité et lexicographie d'apprentissage en langues-cultures étrangères et maternelles*, 135, Paris, Didier-Érudition/Klincksieck, pp. 347-359.
- Silva (R.), 2005 : « Dynamique dénominative et productivité morphologique en Imagerie médicale » dans *De la mesure dans les termes*, actes du colloque en hommage à Philippe Thoiron, Université Lumière Lyon II, Lyon (sous presse).

LA TERMONTOGRAPHIE EN CONTEXTE(S)

Rita Temmerman, Koen Kerremans, Veerle Vandervoort

Erasmushogeschool Brussel

Bruxelles, Belgique

rita.temmerman@ehb.be

Résumé : *en montrant comment les dictionnaires destinés à la traduction spécialisée peuvent être rendus plus performants grâce à une ressource termontographique qui soutiendra les processus de la traduction, nous faisons l'illustration de la méthodologie de la termontographie. Nous commençons par une réflexion sur le processus de la traduction et l'importance des contextes. Ensuite nous prenons le dictionnaire de J. Dancette et C. Réthoré (2000) comme exemple d'un nouveau concept de dictionnaire destiné aux traducteurs. Il est démontré comment la méthodologie de la termontographie et le logiciel Termontography Workbench (TW) pourront rendre l'information dans le dictionnaire de Dancette et Réthoré encore plus accessible et utile aux traducteurs. Nous finissons par quelques idées concernant la recherche au sujet des dictionnaires sensibles aux contextes.*

Mots-clés : *terminologie, traduction, termontographie, ontologie, contexte verbal, contexte situationnel, contexte cognitif*

La problématique des contextes en lexicologie, en terminologie et en traduction touche à plusieurs domaines. Au centre se trouve l'impossibilité d'une délimitation suffisante de la notion de *contexte*. Dans le domaine de la théorie de la traduction par exemple, Durieux (1995) fait la distinction entre le contexte verbal, le contexte situationnel et le contexte cognitif. Certains spécialistes en intelligence artificielle ont essayé de réduire le contexte à une formule mathématique (McCarthy 1987). D'autres ont préféré représenter la connaissance de la situation dans un système qui sert par exemple à faciliter la gestion automatique des dossiers médicaux (Bång *et al.* 1999).

Nous tenterons d'aborder la multiplicité des nuances à l'intérieur de la signification du terme *contextes*, afin de clarifier la particularité de la méthode appelée *termontographie* par rapport aux analyses de la compréhension traditionnellement effectuées par les traducteurs et les terminologues.

Nos recherches et projets en termontographie visant à développer des applications ciblées sur le réseau sémantique et basées sur les ontologies (Temmerman et Kerremans 2003 ; Kerremans 2004) nous fournissent des exemples pour clarifier nos réflexions sur la distinction entre les trois sortes de contextes mentionnées par Durieux.

Un de nos projets de recherche est axé sur les termes de fonctions du domaine de la distribution, afin de représenter l'information essentielle et nécessaire dans une ressource termi-

nologique destinée aux traducteurs. Nous partirons de cet exemple pour illustrer un schéma clarifiant dans lequel nous faisons la distinction entre l'importance des contextes verbaux, situationnels (ou extralinguistiques) et cognitifs pour les traducteurs, les terminologies et les termontographes. On montrera comment notre méthode appelée termontographie et notre logiciel *Termontography Workbench* (TW) pourraient nous permettre de produire des ressources destinées à rendre les dictionnaires plus « intelligents », performants et plus sensibles aux contextes.

Dans la première section de notre texte, nous expliquerons la méthodologie de la termontographie. Ensuite nous essaierons de montrer comment les dictionnaires destinés à la traduction spécialisée peuvent être rendus plus performants grâce à une ressource termontographique qui soutiendra les processus de la traduction. Nous commencerons par une réflexion sur le processus de la traduction et l'importance des contextes (section 2). Ensuite nous prendrons le dictionnaire de J. Dancette et C. Réthoré (2000) comme exemple d'un nouveau concept de dictionnaire destiné aux traducteurs. Nous essaierons de démontrer comment la méthodologie de la termontographie et le logiciel TW pourront rendre l'information dans le dictionnaire de Dancette et Réthoré encore plus accessible et utile aux traducteurs (section 3). À la fin, nous présenterons quelques idées concernant la recherche relative aux dictionnaires sensibles aux contextes (section 4).

1. LA TERMONTOGRAPHIE

Le Centre de recherches en langues spécialisées et de communication (CVC : Centrum voor Vaktaal en Communicatie) de Bruxelles (<http://cvc.ehb.be>) s'est donné comme objectif la combinaison des ressources terminologiques et des ressources ontologiques. Alors que traditionnellement, les ressources terminologiques s'adressent aux utilisateurs humains, une *ontologie* est la définition d'un vocabulaire et de ses relations internes recouvrant un champ de connaissances destiné aux ordinateurs. À la différence d'une terminologie ou d'un vocabulaire, une ontologie cherche à représenter le sens des concepts et des relations qui les lient. Traditionnellement, l'ontologie se définit comme une partie de la métaphysique qui s'applique à l'être en tant qu'être, indépendamment de ses déterminations particulières, c'est-à-dire l'étude des propriétés générales de ce qui existe. Dans l'optique de nos recherches, une ontologie est une spécification explicite d'une conceptualisation dans un domaine et correspond à un vocabulaire contrôlé et à la formalisation explicite des relations qui existent entre les différents termes d'une terminologie. Elle est compréhensible pour les experts d'un domaine et se laisse formaliser afin de soutenir le traitement automatique de l'information. À l'intérieur d'une ontologie nous distinguons différents éléments tels que les classes, les relations, les fonctions, les axiomes formels et les instances.

La termontographie est une méthodologie interdisciplinaire basée sur la théorie sociocognitive en terminologie (Temmerman 2000a et 2000b) pour la gestion et la représentation de la connaissance dans des domaines spécifiques. Cette méthodologie a été développée dans le cadre d'un projet européen (5^e Programme-cadre de recherche) FF-Poirot (<http://www.ffpoirot.org>) et a mené au développement d'un logiciel *Termontography Workbench* (TW). À partir d'une analyse des besoins en vue d'une application spécifique, nous développons une première structure de catégories, notions et relations. Cette structure

sert de modèle de base pour extraire des unités de connaissance d'un *corpus* de textes et évoluera progressivement vers un réseau enrichi et plus pointu de relations sémantiques. Les résultats de cette analyse sont enregistrés dans une base de données terminologiques qui peut servir de ressource pour différents types d'applications (Temmerman et Kerremans 2003). L'approche terminographique se veut multidisciplinaire, en combinant les théories et les méthodes de l'analyse terminologique plurilingue de la théorie sociocognitive (Temmerman 2000a et b) et les méthodes et directives de l'ingénierie ontologique. Partant d'un *corpus* de textes en plusieurs langues, nous faisons la distinction entre la modélisation conceptuelle non spécifique par rapport aux langues et une analyse spécifique des unités de compréhension par rapport à chaque langue et sa spécificité culturelle.

Dans le cadre du projet FF Poirot, le logiciel *Terminography Workbench* (TW) a été développé en partant de plusieurs objectifs. D'abord nous avons voulu utiliser les technologies du Web sémantique pour la détection et la prévention de la fraude en montrant le potentiel des ontologies dans ce domaine. Ensuite nous avons travaillé à une ressource plurilingue terminologique et formelle dans le domaine de la prévention de la fraude financière. Nous avons proposé à la fois des méthodes et des directives pour l'analyse terminologique et la gestion de la connaissance. En parallèle, nous avons développé de nouveaux outils et amélioré des outils existants qui peuvent servir de soutien à l'analyse terminologique et à l'ingénierie de la connaissance.

Afin de comprendre comment la terminographie pourra servir à la production de dictionnaires semi-intelligents pour les traducteurs, nous introduirons des réflexions sur l'importance des contextes dans le processus de la traduction. Sur la base de ces réflexions, nous tenterons de démontrer l'importance des contextes pour la terminographie, qui unira les trois sortes de contextes distingués dans son processus d'analyse.

2. LE PROCESSUS DE TRADUCTION ET LES CONTEXTES

Dans le processus de traduction, le traducteur établit des équivalences entre des unités de traduction dans la langue source et la langue cible. Le processus se fait en deux étapes : le décodage de la signification du texte source et le réencodage de la signification dans la langue cible. Derrière cette procédure simple, il y a une opération cognitive complexe. Afin de décoder la signification du texte source, le traducteur interprète et analyse toutes les caractéristiques du texte, un processus réclamant une connaissance approfondie de la grammaire, de la sémantique, de la syntaxe, des aspects idiomatiques, etc. de la langue source ainsi que la connaissance approfondie de la culture de ses utilisateurs et la compréhension du sujet. Le traducteur a besoin de cette même connaissance afin de réencoder la signification dans la langue cible. Ceci implique l'intuition du traducteur par rapport aux contextes. Nous prenons comme hypothèse qu'un traducteur (humain, système informatique, traduction assistée par ordinateur) ayant accès à des ressources de connaissance terminologique lui procurant de l'information contextuelle produira des traductions de haute qualité. Le défi est dès lors de déterminer le mode de production de ce type de ressources.

2.1. Le traducteur et les contextes

Durieux (1995) a fait la distinction entre les contextes verbaux, situationnels (culturels et psychologiques) et cognitifs de l'unité de traduction. Le traducteur utilise les contextes pour trouver la traduction d'une unité linguistique. Le contexte verbal est constitué par les « unités linguistiques qui précèdent et qui suivent une unité déterminée » (Durieux 1995 : 217). Partant d'un *corpus* en langue de départ et d'un *corpus* en langue d'arrivée, le traducteur saura vérifier l'importance du contexte verbal. Par contexte situationnel, Durieux se réfère à l'« ensemble des données communes à l'émetteur et au récepteur sur la situation culturelle et psychologique [...] de production du texte » (1995 : 217). Le traducteur devra pouvoir estimer où sont les différences culturelles et psychologiques afin de trouver une traduction acceptable. Le contexte cognitif signifie le « stock mnésique qui se constitue au cours de l'assimilation du sens d'un discours ou d'un texte » (Durieux 1995 : 217). Le traducteur se crée le contexte cognitif du texte de départ avant de recréer un contexte parallèle dans la langue d'arrivée.

On montrera comment cette même distinction entre les trois sortes de contextes, faite par Durieux, peut aider à clarifier le travail d'analyse qui devrait être fait par les terminologues produisant des dictionnaires spécialisés de type traditionnel en créant des ressources d'information destinées aux traducteurs (2.2) et comment cette distinction peut être utile au termontographe (2.3) (figure 1).

	le traducteur	le terminologue	le termontographe
Contexte verbal			
Contexte situationnel			
Contexte cognitif			

Figure 1 : Schéma pour la détermination des fonctions comparatives des contextes verbaux, situationnels et cognitifs pour le traducteur, le terminologue et le « termontographe »

2.2. Les terminographes et les contextes

Traditionnellement, le terminologue utilise les contextes verbaux pour savoir décrire et illustrer la signification des termes. Pour lui, le contexte situationnel servira à décrire les différences et variations en une ou plusieurs langues. Le contexte cognitif lui sert dans ce sens qu'il arrivera à formuler une définition basée sur son interprétation de plusieurs contextes verbaux et situationnels.

2.3. Les termontographes et les contextes

Le termontographe utilise ces trois types de contextes introduits par Durieux pour modéliser une partie de la connaissance. Le termontographe se sert des « cotextes » ou de contextes verbaux qu'il trouvera dans un *corpus* de textes pour trouver les relations internationnelles. Le termontographe devrait pouvoir représenter les différences de significations entre deux

ou plusieurs langues et cultures par rapport au schéma de catégorisation. Il collabore avec les ingénieurs de la connaissance qui sont à la base des contextes cognitifs en fournissant des schémas de catégorisation.

Maintenant que nous avons défini l'importance de l'information comprise dans les trois types de contextes, on montrera comment le *Termtography Workbench* (TW) nous a servi dans un projet sur des termes exprimant des fonctions (par ex. gérant de rayon, détaillant, agent d'approvisionnement, commissionnaire d'achats, etc.) dans le domaine de la distribution en partant d'un dictionnaire spécialisé récent.

3. L'INFORMATION CONTEXTUELLE ET LES RESSOURCES TERMINOLOGIQUES

Un bon exemple de dictionnaire bilingue destiné aux traducteurs spécialisés et basé sur l'analyse contextuelle de la compréhension est le *Dictionnaire analytique de la distribution. Analytical Dictionary of Retailing*, de Dancette et Réthoré (2000). Dancette (1998) discute les implications de l'intégration en lexicographie et en terminologie de la dimension cognitive grâce au potentiel qu'il y a à rendre les contextes accessibles dans le dictionnaire informatisé et accessible par internet. Dancette constate que les dictionnaires s'inscrivent aujourd'hui dans une nouvelle réalité pragmatique, théorique et technique. Elle constate une jonction dans les recherches en lexicologie et en traductologie entre les connaissances encyclopédiques et linguistiques. Cette jonction se traduit sur le plan pratique de la rédaction des dictionnaires par le souci de relier les notions et les termes à la multitude des contextes d'utilisation. Pour Dancette, l'évolution dans le développement des dictionnaires spécialisés pour traducteurs va dans le sens de l'élargissement des bases de connaissances et de l'explicitation des relations sémantiques dans les dictionnaires. Elle croit que l'élargissement du contenu tant linguistique qu'encyclopédique des dictionnaires spécialisés bilingues (ou multilingues) s'inscrit dans un contexte technologique qui le rend possible. Pour elle, aussi, grâce à l'exploitation de très larges *corpus* de textes et grâce à ce qu'elle appelle le contexte discursif, la dictionnaire commence à sentir la nécessité de présenter une description des notions qui englobe à la fois la définition sémantique et l'ensemble des réseaux conceptuels qui relient la notion à la multitude de contextes d'utilisation. C'est en ce sens, dit-elle, que le dictionnaire envisagé peut devenir l'interface entre documentation, terminologie et traduction. Son *Dictionnaire analytique de la distribution. Analytical dictionary of retailing* (2000), que nous présentons ci-dessous a été conçu dans cette optique. Par son contenu notionnel hautement spécialisé, d'une part, et la profondeur et la hiérarchisation de l'information, d'autre part, il cherche à répondre aux besoins multiples des traducteurs, rédacteurs, apprenants de la langue spécialisée et professionnel de la distribution. Les nouvelles possibilités technologiques posent la question de la profondeur des informations à inclure dans le dictionnaire, de leur structuration et de leur formalisation.

Ce dictionnaire a été conçu tout d'abord pour les traducteurs traduisant de l'anglais vers le français dans le domaine de la distribution. Leur façon de représenter l'analyse terminologique du domaine prouve l'expérience de Dancette et Rethoré dans les domaines de la didactique et de la recherche en traductologie et en terminologie. À la base de leur approche

terminologique, il y a la conviction qu'un traducteur francophone bénéficiera d'une immersion dans l'information ontologique, notamment de l'information sur la manière dont les termes à traduire sont reliés à d'autres termes dans le même champ lexical ou dans le réseau sémantique de termes liés en français (exemple d'un article, voir figure 2).

L1

LABEL

tag
étiquette

DÉFINITION

Document d'identification du produit qui lui est apposé ou y est attaché et qui en décrit les caractéristiques (nature, prix, provenance, marque, etc.).

PRÉCISIONS SÉMANTIQUES

Les informations contenues sur l'étiquette sont destinées au caissier (prix, codification) et au consommateur (ingrédients, mode d'emploi, etc.). L'étiquette peut présenter un graphisme plus ou moins élaboré.

Depuis les années 1970, l'étiquette comprend généralement un **code-barres** (⇒ BAR CODE). Le code-barres contient des informations telles que la description et le prix du produit, qui seront lues à l'aide d'un **lecteur optique** (⇒ OPTICAL READER).

On distingue l'étiquette **adhésive** (*self-sticking label*), ou **étiquette autocollante**, et l'**étiquette mobile** (*tag*) attachée au produit.

Sur l'**étiquette de prix** (*price tag*) ne figure que le montant à payer.

Aujourd'hui, certaines grandes surfaces n'indiquent plus le prix sur le produit mais plutôt sur la **gondole** (⇒ GONDOLA), avec des **étiquettes de gondole** (*shelf labels, shelf edge labels*) ou des **étiquettes électroniques de gondole** (⇒ ELECTRONIC SHELF LABEL).

RELATIONS INTERNATIONNELLES

Le terme anglais TAG désigne une **étiquette** que l'on peut facilement enlever, ce qui n'est pas le cas de LABEL.

Ne pas confondre l'anglais LABEL avec son homonyme *label*, qui a le sens de **marque** (⇒ BRAND), comme dans le terme PRIVATE LABEL (**marque de distributeur**).

Le terme anglais *quality label* (**label de qualité**), ou *seal of quality*, ne désigne pas l'étiquette

mais la conformité du produit à des caractéristiques ou à des qualités définies.

COMPLÉMENTS D'INFORMATION

L'idée de remplacer le **marquage** (*marking*) des prix par des codes-barres sur les produits vendus en magasin a suscité beaucoup de craintes chez les consommateurs. Ils y voyaient un préjudice à l'exercice de leur droit à l'information. Pour répondre à cette inquiétude certains distributeurs mettent des **lecteurs de codes-barres** (⇒ BAR-CODE READER) à la disposition de leurs clients, dans les divers rayons de leurs magasins. Par contre, il est admis que le code-barres présente des avantages par rapport à l'étiquette de prix, dont l'exactitude et la rapidité des enregistrements à la caisse.

Les producteurs ont l'obligation, en vertu de la **Loi sur la protection du consommateur** (*Consumer Protection Act*), de répertorier sur l'étiquette tous les ingrédients contenus dans le produit alimentaire.

INFORMATIONS LINGUISTIQUES

Marquage a été normalisé par l'OLF (1983) et l'ISO (1987).

étiqueter: to ticket
étiqueteuse: labeler, label machine

CONTEXTES

But it wasn't until 1900 that [he] put the first Polar label on a bottle of cool, naturally purified water taken directly from one of these springs on his property.

<<http://www.water.com/polar/index.html>> (30-3-99)

Dans ce but, la réglementation mise au point par les organismes de la CEE et par l'administration française prévoit sur chaque étiquette la présence d'un certain nombre de mentions obligatoires, en fonction de la catégorie du vin.

<<http://www.vin-champagne.com/etiq.htm>> (30-3-99)

LADS Voir LIMITED ASSORTMENT DISCOUNT STORE

Figure 2 : Exemple d'un article dans le dictionnaire de Dancette et Réthoré

Toute cette information est donnée dans la langue cible (le français) afin de stimuler l'autonomie discursive du traducteur au moment de la formulation du texte en langue cible (pour un compte rendu plus approfondi, voir Temmerman 2003). L'explicitation des « relations internationnelles » prend la forme de définitions intensionnelles et extensionnelles. L'utilisateur du dictionnaire trouve l'information sur la signification d'un terme anglais en plusieurs modules textuels formulés en français : la définition, les précisions sémantiques, les relations internationnelles, les compléments d'information, les informations linguistiques, les contextes verbaux et les exemples. Dans les textes, les vedettes faisant l'objet d'un autre article sont marquées ainsi que les termes mis en relations. Dancette et Réthoré déterminent leur propre stratégie de définition des termes en se basant sur l'analyse des besoins de l'utilisateur potentiel du dictionnaire. La définition traditionnelle d'un terme générique, complétée de *differentiæ* est enrichie d'informations sémantiques supplémentaires, d'informations internationnelles, d'informations linguistiques et de contextes explicatifs. La prototypicité de la signification des termes et des catégories est prise en considération, et les termes métaphoriques basés sur des analogies sous-jacentes sont traités de manière conséquente ainsi que la polysémie et la synonymie. En partant d'une perspective sémasiologique, ils arrivent à décrire les termes attestés dans un *corpus* de textes spécialisés en tenant compte de la fréquence d'occurrences. Ils réussissent à rajouter un maximum d'informations sur les liens relationnels (l'information ontologique) dans la partie descriptive de chaque article. De tels dictionnaires ouvrent des perspectives pour les projets lexicographiques et terminographiques de l'avenir. Ils pourront devenir encore plus faciles à l'emploi ainsi que plus performants si l'on utilise des outils semi-intelligents soutenus par des bases de connaissances comme des ontologies.

Dans notre projet pilote portant sur les termes des fonctions dans le domaine de la distribution en anglais et en néerlandais, nous sommes partis de termes figurant dans le dictionnaire de Dancette et Réthoré (*dealer, leasing director, affiliated retailer, merchandiser, rack jobber, detailer*, etc.). Nous avons voulu tester comment notre méthode terminographique, en combinaison avec des outils de gérance de la connaissance, peuvent produire des banques de données terminologiques plus « intelligentes », grâce au fait que l'information contextuelle est représentée de manière plus complète. Nous sommes partis d'un *corpus* de textes concernant la distribution en néerlandais et en anglais (*contextes verbaux*). Les *contextes situationnels* se situent au niveau des différences culturelles et les *contextes cognitifs* au niveau d'un schéma de catégorisation qui servira de source d'inspiration pour la définition des « classes » dans le *Terminography Workbench* (TW), le logiciel de gestion en terminographie développé par CVC (figure 3).

Le TW se constitue de plusieurs fenêtres de travail (figure 3) :

- la fenêtre du schéma de la catégorisation ;
- la fenêtre de la liste des termes montrant tous les termes sélectionnés par rapport à chaque concept faisant partie du schéma de la catégorisation (figure 4) ;
- la fenêtre de textes (figure 3 milieu : où les textes du *corpus* peuvent être chargés et où l'analyse terminologique peut être exécutée) ;

- la fenêtre de l'annotation des sources (figure 3 à droite) où les résultats de l'analyse sont visualisés. Nous voyons par exemple (figure 3) le terme qui a été sélectionné (T) et une description (D) ;
- la fenêtre dans laquelle nous ajoutons une description (figure 5) ;

Le TW nous permettra d'intégrer les trois types de contextes (verbaux, situationnels et cognitifs) dans les ressources termontologiques.

- Les contextes verbaux intéressants, comme les définitions que nous retrouvons dans les textes du *corpus* de l'analyse, peuvent être liés aux termes dans la fenêtre destinée aux descriptions (figure 5).
- Les contextes situationnels : les différences de signification entre des termes mis en relation peuvent être explicitées dans une fenêtre destinée aux commentaires de transfert. Les commentaires de transfert s'adressent aux utilisateurs de la ressource (dans ce cas-ci, les traducteurs) et sont prévus dans le but d'informer l'utilisateur des dangers d'erreurs de transfert d'information d'un contexte situationnel ou culturel à l'autre, c'est-à-dire d'un terme source à un terme cible. Pour l'instant la fenêtre « commentaires de transfert » est un champ textuel. Nous avons l'intention de tester d'autres possibilités.
- Les contextes cognitifs : la terminologie trouvée dans les textes du *corpus* peut être liée aux concepts du schéma de la catégorisation. L'exemple en figure 4 montre comment le terme néerlandais *winkelbediende* trouvé dans le texte (voir figure 3, fenêtre du milieu) est lié au concept symbolisé par le terme anglais *salesperson*.

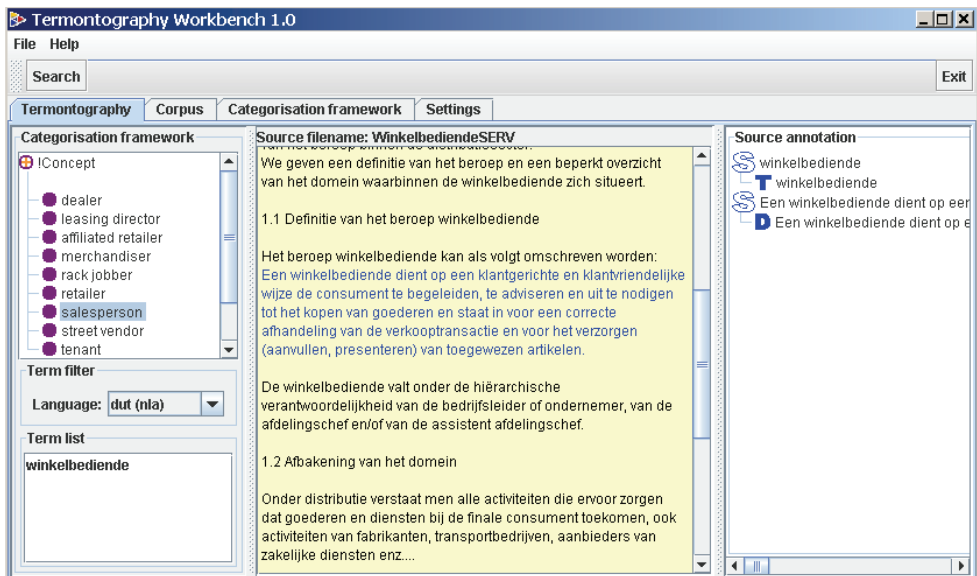


Figure 3. Le *Termontography Workbench*

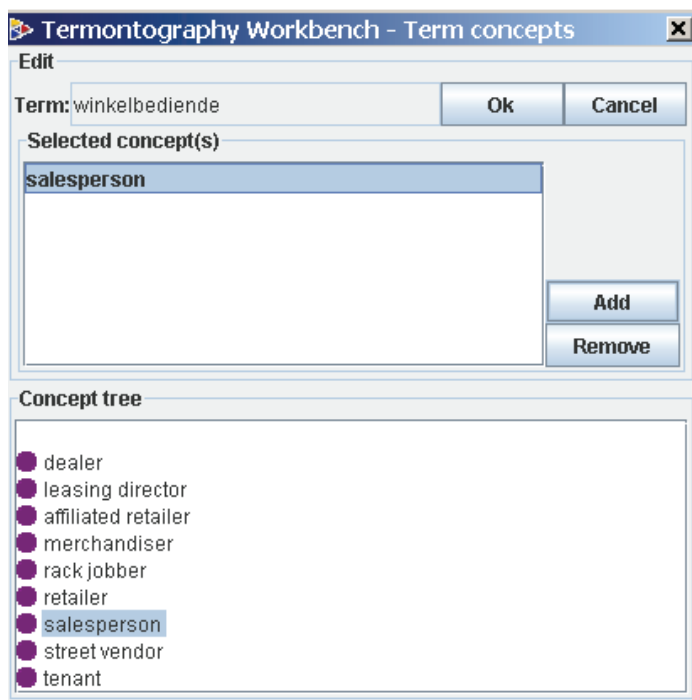


Figure 4. Fenêtre des termes liés à un concept



Figure 5. Fenêtre dans laquelle il est possible d'ajouter une description

4. À LA RECHERCHE DES DICTIONNAIRES SENSIBLES AUX CONTEXTES

Dancette et L'Homme (2004) expliquent leur tentative de convertir le dictionnaire de Dancette et Réthoré en base de données relationnelle en utilisant les fonctions lexicales selon

Mel'čuk. Notre approche se dirige plutôt vers l'analyse de l'information (concepts, termes et relations) obtenue dans trois contextes différents.

Les problèmes à résoudre pour tenter de produire des ressources terminologiques automatisées pour les traducteurs restent entre autres les suivants :

- à part la définition essentielle, quelle quantité d'information contextuelle la ressource terminologique devrait-elle procurer ?
- comment cette information contextuelle doit-elle être structurée ?
- à quel point cette information contextuelle peut-elle être formalisée dans l'optique de soutenir le processus de traduction ?

Pour ce qui est de la fonctionnalité, nous sommes à la recherche d'un système intelligent s'adaptant de manière dynamique au profil de l'utilisateur, ou autrement dit, nous sommes à la recherche d'un dictionnaire automatique pluri-accessible s'adaptant à l'utilisateur. Nous devons faire la distinction entre la sensibilité aux contextes du point de vue des traducteurs (les ressources contenant de l'information sur plusieurs types de contextes) et la sensibilité aux contextes du point de vue des ressources, dans le but de créer des ressources s'adaptant aux utilisateurs. Nos recherches au sujet des dictionnaires sensibles aux contextes en relation avec le développement en cours du *Termontography Workbench* s'inscrivent aussi dans l'intérêt qu'on a à retrouver quels éléments du contexte d'utilisation d'une ressource remettent en question la pertinence de la ressource. Nos recherches évolutives sur la portée de la distinction entre les contextes verbaux, situationnels et cognitifs et sur l'application de cette distinction dans l'analyse terminologique partant des *corpus* devront nous mener à une adéquation plus pointue d'une ressource dans de nouvelles circonstances telles que l'élargissement du *corpus*, de nouveaux besoins ou des altérations dans le domaine.

BIBLIOGRAPHIE

- Bång (M.), Eriksson (H.), Lindqvist (K.) et Timpka (T.), 1999 : « An Approach to Context-Sensitive Medical Applications » dans *Proceedings of the 1999 AMIA Annual Fall Symposium*, Washington D.C.
- Dancette (J.), 1998 : « Le potentiel du dictionnaire spécialisé bilingue électronique » dans *actes sélectifs – Euralex '98 Proceedings*. Université de Liège.
- Dancette (J.) et Réthoré (C.), 2000 : *Dictionnaire analytique de la distribution. Analytical dictionary of retailing*, Les Presses de l'Université de Montréal.
- Dancette (J.) et L'Homme (M.-C.), 2005 « Building specialized dictionaries using lexical functions » dans Temmerman (R.) et Knops (U.), édit. *The Translation of Domain Specific Languages and Multilingual Terminology Management*, numéro spécial de *Linguistica Antverpiensia*, pp. 113-132.
- Durieux (C.), 1995 : « Texte, Contexte, Hypertexte » *Cahiers du Ciel 1994-1995*.
- Kerremans (K.), 2004 : « Categorisation frameworks in termontography » dans Temmerman (R.) et Knops (U.), édit., *The Translation of Domain Specific Languages and Multilingual Terminology Management*, numéro spécial de *Linguistica Antverpiensia*, pp. 263-278.
- McCarthy (J.), 1987 : « Generality in Artificial Intelligence » *Communications of the ACM*, vol. XXX, n° 12, pp. 1030-1035.

- Temmerman (R.), 2000a : *Towards New Ways in Terminology Description. The Sociocognitive Approach*, Amsterdam/Philadelphia, John Benjamins.
- Temmerman (R.), 2000b : « Une théorie réaliste de la terminologie : le sociocognitivism » dans *Terminologie Nouvelles* 21, pp. 58-64.
- Temmerman (R.), 2003 « Innovative Methods in Specialised Lexicography » dans *Terminology* 9 (1), pp. 117-135.
- Temmerman (R.) et Kerremans (K.), 2003 : « Termontography : Ontology Building and the Sociocognitive Approach to Terminology Description » dans *Proceedings of CIL17*, Prague.

LA TRADUCTION DE LA MÉTALANGUE : LA PROBLÉMATIQUE *TERME/MOT* EN CONTEXTE

Béchir Ouerhani

Université du 7 novembre à Carthage
Institut supérieur des langues appliquées et d'informatique de Nabeul
Tunis, Tunisie

***Résumé :** L'une des spécificités fondamentales de la terminologie linguistique est qu'elle puise ses sources dans les mots de la langue, d'où sa dimension autonymique. Le résultat est que les emplois de la terminologie vacillent selon les contextes entre terme et mot. Il s'ensuit qu'une définition rigoureuse du contexte est exigée pour que la définition de l'unité soit mieux cernée. Cette problématique conditionne la qualité de la traduction de ces textes d'une langue en une autre. Si on n'en tient pas compte, les risques de confusions sont énormes et la qualité de la traduction s'en ressent. Nous partirons d'exemples précis empruntés à certaines théories et à des textes traduits du français vers l'arabe pour illustrer notre problématique.*

***Mots-clés :** terme, contexte, traduction, terme linguistique*

1. INTRODUCTION

Notre exposé partira du fait suivant : l'une des spécificités fondamentales de la terminologie linguistique est qu'elle puise ses sources dans les mots de la langue. Le résultat est que les emplois de la terminologie vacillent selon les contextes¹ entre terme et mot.

Par ailleurs, il est possible de répartir les termes linguistiques en deux groupes selon leur « degré de spécialisation », c'est-à-dire le degré d'appropriation au domaine d'utilisation. Nous allons essayer de montrer que les difficultés que rencontre le traducteur de la terminologie linguistique – à travers la traduction des textes – varient selon ce « degré de spécialisation ».

Notre propos s'appuiera sur l'analyse d'exemples précis empruntés à des textes traduits du français vers l'arabe et de l'arabe vers le français.

2. UN PREMIER GROUPE : LES TERMES « HYPERSPÉCIALISÉS »

L'examen de ces termes, aussi bien dans l'emploi au sein d'une langue donnée que lors de leur traduction, montre qu'ils ne posent pas de problèmes particuliers puisque les domaines de leur emploi sont plutôt restreints.

¹ Nous observons que nous utilisons ce terme dans son acception la plus générale et indépendamment du débat – parfois vif – sur la définition de ce qu'on entend par « contexte » (Cf., par exemple certaines communications présentées ici-même). En ce qui concerne notre centre d'intérêt, le contexte va de l'environnement de droite et gauche d'un terme donné jusqu'au type de discours, puisqu'ils participent tous à la détermination du statut du terme en question et déterminent, par la suite, notre choix pour la traduction.

2. 1. Au sein de la même langue :

L'emploi de ce type de terme est lié à un contexte bien déterminé : ce sont des termes qui impliquent leur définition à travers l'implication d'un cadre théorique, d'une école ou d'un linguiste. En effet, les contextes dans lesquels ces termes sont utilisés sont extrêmement marqués. Une fois utilisés, ces termes impliquent tout de suite des textes de référence ou les cadres théoriques desquels ils sont issus. On remarque ce phénomène aussi bien dans les textes linguistiques que dans les dictionnaires de linguistique. Mais c'est dans ces derniers que ce fait est plus marquant vu le caractère métalinguistique du contenu de ce type de textes. Tel est le cas, par exemple de la plupart des termes guillaumiens ou certains termes fortement liés à l'école générativiste ou transformationnelle. Ainsi, pouvons nous lire dans le *Dictionnaire des sciences du langage* de F. Neveu (2004) pour les termes *chronotype*, *monème* et *structure profonde*, à titre d'exemple :

[1] Chronotype : « CHRONOTYPE [psychomécanique] [...] La notion de chronotype est développée par Gustave Guillaume à partir de 1929 dans *Temps et Verbe*. Elle désigne chacune des deux parcelles de temps que prélève le présent pour sa constitution, notamment en français... ».

Monème : « MONÈME [...] La notion de monème a été utilisée par André Martinet pour désigner les unités significatives minimales, qui constituent la première articulation du langage ».

Structure profonde : « STRUCTURE PROFONDE Dans le cadre théorique de la grammaire générative et transformationnelle, le domaine syntaxique est organisé en deux composantes [...] ».

Dans ces trois passages, on voit clairement que la définition de ces termes est tout de suite liée au cadre théorique dans lesquels ils s'inscrivent. Cela suffit pour lever toute ambiguïté du côté du lecteur, puisque seulement deux cas de figure sont possibles : soit il a une certaine idée sur le cadre théorique évoqué, soit il n'en a pas. Dans les deux cas, le lecteur n'est pas confronté au problème de la détermination du statut du terme : il n'a pas à trancher s'il s'agit d'un emploi de la langue courante (il est donc en présence d'un mot) ou d'un emploi spécialisé (il a affaire, dans ce cas, à un terme), puisque le statut est tout de suite rappelé par l'évocation du cadre théorique ou de l'auteur.

Dans les textes linguistiques, soit l'auteur s'inscrit de façon explicite dans le cadre théorique qu'implique le terme en question, soit il adhère à certaines analyses proposées. Dans les deux cas, on y trouve souvent des renvois aux définitions ou aux textes de référence concernant telle ou telle théorie. Notons que même en cas d'absence de ces renvois, ces termes impliquent chez le lecteur leurs propres définitions vu leur « haut degré de spécialisation ».

Sur le plan sémantique, nous pouvons évoquer la question de la transparence de ces termes. En effet, ces termes hyperspécialisés ne posent pas généralement de problème puisqu'ils réfèrent chacun à un concept bien déterminé et n'ont pas d'emploi dans la langue courante. Ils sont très proches alors du principe de la monosémie exigé pour les termes techniques, très cher à la littérature sur les langues de spécialité (Voir, entre autres Lerat 1995 : Chap. VIII).

Qu'en est-il alors de leur traduction ?

2. 2. Sous l'angle de la traduction :

Pour les raisons évoquées dans le paragraphe précédent, la traduction de ce type de termes ne pose pas de problèmes particuliers. Nous pourrions affirmer simplement que leur transfert est toujours lié au contexte qu'ils évoquent et aux cadres théoriques qu'ils impliquent. Si l'on prend, par exemple, l'expérience de notre équipe de traduction, nous constatons que les difficultés de la traduction de ces termes hyperspécialisés concernent plutôt :

- la formation du terme (si l'équivalent existe ou s'il faut recourir à la néologie). Nous rappelons à ce propos que pour une langue comme l'arabe, nous sommes tout le temps confronté à cette question vu l'énorme quantité de savoir à transférer, et par conséquent le nombre important de concepts nouveaux et de termes qui les désignent.
- l'harmonisation du terme proposé avec le paradigme des termes traduits et avec la grille des termes liés au terme en question par un lien quelconque (même cadre théorique, spécification, niveau de l'analyse, etc.).
- la prise en compte des paradigmes de terminologie linguistique déjà existants, nous entendons par là notre système terminologique hérité d'une grammaire arabe ancienne de plus de quatorze siècles. En effet, la tradition grammaticale dispose souvent d'un terme qui rend compte, totalement ou partiellement, de la notion en question. Dans d'autre cas, il est possible de prendre un terme déjà existant et d'en modifier le contenu. (Pour certains aspects de ces questions, nous renvoyons à Neveu et Ouerhani, à paraître).

En voici quelques exemples : les termes guillaumien *cinétisme* et *cinèse* sont traduits par des néologismes :

[2]

Cinétisme	<i>ħara : kijja</i> حَرَائِيَّة
Cinèse	<i>ħara : ka</i> حَرَائِيَّة

et cela ne nécessite pas une précision particulière quant au contexte puisqu'il est automatiquement impliqué par le simple fait de les employer.

Quant au terme *Monème*, il a été traduit depuis très longtemps par le terme لفظم *lafðam* qui est une composition du mot *lafð* (l'équivalent de « terme » dans la langue courante) et le suffixe *Amé*, traduction du suffixe « ème » :

[3] *Monème* : لفظم (*lafðam*)

Vu le caractère monosémique de ces termes, le traducteur n'est pas confronté au problème de la dualité mot/terme.

Pour conclure ce premier point, nous nous permettons d'insister sur le fait que ces termes impliquent tout de suite lors de leur emploi leurs contextes : en un mot, quand on les traduit, ils sont transférés avec leurs contextes d'emploi et avec leurs implications propres (définition, cadre théorique, historique, etc.). Il s'ensuit que la dualité mot/terme ne se pose pas à leur égard.

Examinons à présent le deuxième groupe défini au début de notre exposé.

3. UN DEUXIÈME GROUPE : LES TERMES PUISÉS DANS LA LANGUE COURANTE

Ces termes sont employés par ailleurs en tant que mots de la langue courante. Vu leur appartenance initiale au lexique général d'une langue donnée, ils sont généralement polysémiques. Ce qui revient à dire qu'ils ont plus d'un emploi et que les différentes significations auxquelles ils renvoient sont spécifiées selon le contexte d'emploi. De ce fait, leur statut de terme spécialisé ou de mot de la langue courante est déterminé par le contexte. Nous pensons à la grande majorité des termes linguistiques qui viennent de la langue courante. Cela semble concerner toutes les langues naturelles. En voici quelques exemples du français et de l'arabe :

(4)

Mot	<i>kalima</i> كَلِمَة
Lettre	<i>ḥarf</i> حَرْف
Phrase	<i>ḡumla</i> جُمْلَة
Nom	<i>ḡism</i> إِسْم
Adjectif	<i>ḡifa</i> صِفَة
Etc.	إلخ.

Prenons, à titre d'exemple, le terme arabe *ḡifa* (5^e dans le tableau ci-dessus). Ce terme est employé en arabe dans le sens de :

« qualité » quand il s'agit d'attribuer une qualité à quelqu'un :

لزید صفات حسنة]5]

Li-zajd-sifa : t-ḥasana

A- Zayd- qualités *bonnes*

Zayd a de bonnes qualités

« adjectif » quand il s'agit d'analyse ou de terminologie linguistiques :

["مريض" صفة إسنادية في زيد مريض]6

"mari:d"-sifa-ḥisna:dijja-fi:-zajd-mari:d

« malade »- adjectif- prédicatif- dans- *Zayd- malade*

« malade » est un adjectif prédicatif dans *Zayd est malade*

Si l'on applique l'analyse selon le schéma d'arguments, nous constatons que *-sifa:t* dans [5] est un prédicat dont le premier argument est le nom propre « Zayd ». Il est par ailleurs en position d'argument d'un prédicat de rang supérieur : *ḥasana*. Cet « environnement » impose donc l'attribution du statut de « mot » de la langue courante au terme en question.

Dans [6], il s'agit d'une métalangue comme l'indique la présence de guillemets, des termes *sifa-ḥisna:dijja* attribués au terme en question, et, enfin, la présence d'un exemple sur lequel porte ce commentaire. Il s'agit donc d'un emploi spécialisé du terme *sifa*.

Sur le plan conceptuel, nous sommes en présence de deux emplois différents. En effet, si l'emploi courant renvoie à une signification, il y a derrière l'emploi terminologique un et un seul concept qui se situe dans tout un réseau de relations entre concepts appartenant à la même discipline.

Sous l'angle de la traduction, la détermination rigoureuse du contexte est nécessaire pour décider du statut (mot/terme) de l'élément en question. Pour le traducteur, ce type de termes pose donc plus de problèmes que ceux du premier groupe.

Dans ce qui suit, nous allons examiner de près certains cas de ces termes qui vacillent entre mot et terme selon le contexte. Nous allons voir que la détermination du contexte d'emploi (courant ou spécialisé, ou même selon les spécialités) est nécessaire pour garantir la réussite à l'opération de transfert.

3. 1. Du français vers l'arabe :

3. 1. 1. « terme » et « terme »

Selon qu'il s'agit de langue de spécialité ou de langue courante, le mot, ou justement le terme *terme*, peut renvoyer soit à une unité appartenant au « lexique général », soit à une unité appartenant à un vocabulaire technique spécifique à une discipline. Mais si nous quittons déjà la langue courante pour examiner de près son emploi dans les textes linguistiques, nous remarquons que les deux emplois existent. Nous avons choisi d'illustrer cela à partir de l'ouvrage de G. Gross et M. Prandi et de sa traduction arabe sur laquelle notre équipe est en train de travailler :

[7] La finalité (p. 4) : « Chacun des termes prédicatifs appartenant aux trois aires, en outre, peut aussi bien entrer dans la construction de la phrase complexe, incorporé dans l'une de ces structures que la tradition appelle locutions prépositionnelles [...], qu'orienter l'interprétation d'une juxtaposition en tant qu'expression anaphorique ».

En examinant de près ce passage, nous observons que le terme « terme » est utilisé ici dans le sens d'« unité du lexique général » et non pas dans le contexte de langue spécialisé. En effet, les auteurs parlent des unités lexicales prédicatives d'une manière générale et il ne s'agit nullement de termes spécifiques à une discipline quelconque. C'est pour cela que nous l'avons traduit en arabe par لفظ *lafḥ* (l'équivalent de « terme » dans la langue courante) que nous avons vu plus haut :

"يمكن لكل واحد من الألفاظ الإسنادية المنتمية إلى الفضاءات الثلاثة، علاوة على ذلك، أن يدخل في تركيب جملة مركبة، مندمجاً في إحدى البنى التي يسميها التقليد عبارات حرفية (...) أو أن يوجه تأويل التجميع من حيث هو عبارة ترددية".

Jumkinu-li-kulli-wa:hidin-min-al-ʔalfa:ḥ... (ʔalfa:ḥ=pluriel de *lafḥ*)

Plus loin, au début du premier chapitre, les auteurs disent : « Nous utilisons le terme de relation transphrastique pour référer à des relations entre procès comme la cause, la concession, le but ou la conséquence » (p. 13).

Il s'agit là d'une unité lexicale utilisée dans un domaine spécifique pour renvoyer à un concept donné, celui des relations entre procès. Ceci est renforcé par la présence :

- du prédicat « utiliser » dont l'argument est le terme « terme » ;
- de l'expression du but (introduite par « pour ») et qui contient le verbe « référer à » dont l'un des arguments est « les relations entre procès » ;

nous sommes donc en présence d'un terme renvoyant à un concept dans le domaine de la linguistique. Ce qui détermine notre choix pour la traduction : le terme *muṣṭalaḥ* (équivalent de *terme* dans le domaine des langues de spécialité) :

"نحن نستعمل مصطلح علاقة بين-جمليّة للإحالة على العلاقات بين الأعمال مثل السبب أو الإضراب أو الهدف أو النتيجة"

naḥnu-nastaʕmilu-muṣṭalaḥa-ʕala:qa-bajn-ʕumlijja...

Les exemples concernant la problématique de la dualité mot/terme dans la traduction de textes linguistiques en arabe sont multiples. Pour ne citer que quelques-uns, nous renvoyons à la traduction des cours de F. de Saussure réalisée par S. Garmadi, M. Chaouech et M. Ajina (1985), l'article de L. Guilbert « La néologie », *GLLF* et sa traduction arabe réalisée par T. Baccouche et S. Mejri (1993), ainsi que la traduction des *Clefs pour la linguistique* par T. Baccouche (1993).

3. 1. 2. Le terme proposition entre langue courante et langues de spécialité

Selon les contextes, ce terme a en français un emploi courant (lié à la construction *faire une proposition*) et plus d'un emploi spécialisé selon la discipline, et donc selon le réseau conceptuel auquel appartient le concept qui réside derrière le terme. Ainsi, il y a une différence d'emploi entre la *proposition* dans le domaine de la logique et la *proposition* dans le

cadre de l'analyse de la phrase complexe, par exemple, même si on pourrait affirmer l'existence d'un certain lien entre les deux concepts. Cette différence est bien explicite quand il s'agit de traduire le terme vers une langue comme l'arabe. Il suffit de regarder, par exemple, dans les dictionnaires de langue bilingues et dans les dictionnaires de linguistique bilingues ou les glossaires qui accompagnent les ouvrages traduits et les travaux de recherche. En voici quelques échantillons :

[8]

– *Al-Manhal* (dictionnaire bilingue français/arabe) :

Proposition sf.

	تقديم، عرض، مقترح، اقتراح (...)
<i>taqdi:m, ʕarḍ, ʔiqṭira:ḥ, muqtaraḥ</i>	
<i>Faire des_s à une femme</i>	راود امرأة
<i>ra:wada-mraʔatan</i>	
<i>_(Maths.)</i>	اقتراض
<i>ʔiftira:ḍ (litt. Hypothèse)</i>	
<i>_(Log.)</i>	قضیة
<i>qaḍija (litt. Affaire)</i>	
<i>_(Gramm.)</i>	جملة
<i>ʕumla (phrase)</i>	

Notons que ce dernier emploi (*Gramm.*) est plutôt traduit chez les linguistes arabes par :

- *qaw*/en linguistique générale et dans le contexte des grammaires logiques.

- *ʒumla-soʒra*: dans un contexte concernant la phrase complexe, ce qui veut dire littéralement petite phrase par opposition à la *ʒumla* (la principale) (cf., à titre d'exemple M'seddi 1984).

Observons, par ailleurs, que dans le cadre de notre équipe de traduction, le terme linguistique *proposition* est systématiquement traduit par : *qawl*. Dans le contexte de la logique et notamment des syllogismes, il est traduit par *qadlǰja*.

Déterminer le contexte avec précision est donc indispensable pour réussir la traduction de ce terme en arabe. Examinons à présent le cas du passage de l'arabe vers le français.

3. 2. De l'arabe vers le français

Nous allons voir quelques termes en arabe dont le statut dépend du contexte et dont la traduction par la suite est déterminée par ce statut :

[عِبارة 9] [*ʒiba:ra*]

Littéralement, ce terme signifie « expression », c'est d'ailleurs sa traduction la plus courante en français. Examinons toutefois ces exemples :

['لا تعجبني هذه العبارة 9]

la:-toʒibuni:-ha:ʒihi-l-ʒiba:ra

négation plaire-moi cette-la- expression

Cette expression ne me plaît pas (*langue courante*)

['مفهوم العبارة أساسي في دراسة ظاهرة التكلس 9]

mafhu:mu-l-ʒiba:ra-ʒasa:sujjun-fi:-dira:sati-ḍa:hirati-t-takallusi

notion-la-locution- fondamentale- dans- étude- phénomène-le-figement

La notion de locution est fondamentale dans l'étude du phénomène du figement

Dans le premier exemple, le contexte nous présente un mot de la langue courante, même s'il est utilisé par ailleurs en tant que terme linguistique. En effet, nous déduisons de la présence du prédicat *toʒibu* (« plaire »), dont le terme *ʒiba:ra* est le premier argument et dont le locuteur est le second argument, qu'il s'agit d'un jugement que fait un locuteur « ordinaire » concernant un énoncé, ou une partie d'un énoncé. Le contexte a donc déterminé la traduction par le terme *expression* dans le sens de la langue courante. Tandis que dans le second exemple, le contexte d'emploi est spécifié : nous sommes en présence d'un commentaire linguistique concernant le terme de *ʒiba:ra* qui est utilisé comme tête de paradigme dans des termes composés tels que :

[10]

- locution prépositionnelle	- عبارة حرفية
- locution verbale	- عبارة فعلية
etc.	إلخ.

]11 [تدلّ الحال [Le *ħa:l* (litt. *État*)

Examinons les deux emplois suivants :

['هذا المحرك في حال جيدة]11

*ħaḍa-l-muħarrik-fi-ħa:l-ʒajjida**ce- le- moteur- dans – état- bon**Ce moteur est dans un bon état*

['تدلّ الحال على هيئة الفاعل أو مفعول أو كيفية وقوع الحدث...]11

*tadullu-l-ħa:lu-ʕala:-ħajʔati-l-fa:ʕil-ʔaw-l-maʕnu:l-ʔaw-kajfijjat-wuqu:ʔi-l-ħadaθ...**désigne- le- état- prép.- attitude- le- sujet- ou- le- objet- ou- manière- déroulement- action**Le ħa:l désigne l'état du sujet ou de l'objet, ou la manière du déroulement de l'action (procès)... (Manuel scolaire de l'enseignement de base).*

Dans ces deux phrases, nous sommes en présence de deux emplois différents du terme *ħa:l*. Le premier relève de la langue courante et le terme renvoie à une signification non spécialisée. Quant au second, il appartient à une langue de spécialité, la terminologie grammaticale arabe et renvoie à un concept : celui de l'équivalent de l'adverbe de manière en français. Ce deuxième emploi est précisé par le contexte qui indique clairement qu'il s'agit d'une métalingue.

Notons que ceci n'a pas empêché un certain nombre de confusions chez certains écrivains de grammaires arabes pour étrangers. En fait, certains ont traduit le terme spécialisé *ħa:l* par son emploi courant et ont proposé donc « état » pour le français par exemple.

Observons aussi que le problème de l'emploi et de la traduction du *ħa:l* est doublé par une autre difficulté qui vient s'y ajouter, quand on est dans le domaine de la terminologie grammaticale : on est en fait tenté de le traduire par le terme *adverbe*, ce qui n'est pas perti-

ment vu que le phénomène du *ḥa:l* arabe ne couvre qu'une partie de la notion de l'adverbe en français. Nous nous contentons de rappeler rapidement que l'aire que couvre l'adverbe français est répartie en arabe entre plusieurs phénomènes, dont chacun possède son propre terme (cf. Mosbah S., à paraître ; Ouerhani B., à paraître) :

- *ḥa:l* (litt., État, équivalent de l'adverbe de manière),
- *maḥḥu:l mutlaq* (équivalent de l'objet interne, exprimant souvent l'intensité),
- *tamji:z* (sorte de spécifieur exprimant la qualité ou la quantité par rapport à la prédication ou à l'un de ses éléments).

Le *fa:ʕil* "الفاعل" [12]

Le cas du *fa:ʕil* est un cas typique du double emploi d'un terme (courant et spécialisé), doublé de deux usages selon le contexte, au sein même de la langue de spécialité. Ce qui lui donne à la fin trois emplois distincts.

– dans la langue courante, le mot *fa:ʕil* signifie l'idée générale de « celui qui a fait quelque chose ».

– dans la terminologie grammaticale, il est l'équivalent :

→ de la fonction *sujet* en français.

→ de l'*agent*, à un niveau d'analyse autre que celui des fonctions syntaxiques.

Il faut donc déterminer avec précision le contexte d'emploi pour pouvoir décider par quel terme le traduire. Réciproquement, le problème se pose aussi quand on part dans l'autre sens, c'est-à-dire quand on veut traduire en arabe les termes de *sujet* et d'*agent*, puisque l'arabe ne dispose que d'un seul terme pour renvoyer aux deux concepts. Ainsi, dans la traduction de *Pour une logique de sens* (R. Martin), nous avons été confronté à ce problème. Le choix que nous avons opéré était de garder le même terme *fa:ʕil* pour traduire et le sujet et l'*agent*.

Pour en finir avec ce cas, nous observons que le même terme de *fa:ʕil* est utilisé dans le domaine de la morphologie dérivationnelle arabe pour désigner un schème très productif de l'arabe : celui du nom d'*agent*².

En ce qui concerne notre équipe de traduction, nous avons choisi donc de garder l'unique terme arabe et de compter sur le contexte pour séparer les emplois. Derrière ce choix réside le fait que ce terme, avec tous ses emplois, est très ancien et stabilisé aussi bien dans la langue courante que dans la terminologie grammaticale.

² Nous rappelons que les « schèmes » peuvent être définis comme des combinaisons de consonnes et de voyelles correspondant à des significations différentes et qui servent de moules pour la formation régulière de différents dérivés appartenant au même type à partir de la matière consonantique. C'est la raison pour laquelle la morphologie dérivationnelle de l'arabe est quasi systématique.

3. CONCLUSION

En regardant la grande confusion qui règne actuellement sur la terminologie linguistique arabe, nous ne pouvons qu'insister sur l'importance de la prise en compte du contexte aussi bien dans « le sens large » que dans « le sens étroit » du terme. C'est dans cet esprit que nous proposons que l'élaboration de dictionnaires de linguistique pour l'arabe doit nécessairement passer par la traduction des termes dans leurs contextes d'emploi.

C'est dans cette perspective que notre équipe travaille sur la traduction de textes linguistiques « fondateurs » qui serviront de base à l'élaboration de dictionnaires de linguistique ou à tout travail terminologique. La traduction de certains textes fondateurs de la grammaire arabe en français s'inscrit également dans le même cadre.

BIBLIOGRAPHIE

- Baccouche (T.) et Mejri (S.), 1996 : « La problématique du système de terminologie linguistique arabe. Échantillon méthodologique » dans *Études linguistiques*, 1, Publications de l'ATL, Tunis, pp. 9-23 (en arabe).
- Gross (G.) et Prandi (M.), 2004 : *La Finalité, fondements conceptuels et genèse linguistique*, Duculot.
- Guilbert (L.), 1986 : « La néologie », dans *GLLF*, pp. 3584-3594.
- Idriss (S.), 2002 : *Al-Manhal*, dictionnaire bilingue français/arabe. Beyrouth, 30^e édition.
- Collectif, *Le Petit Robert électronique*, édit., 2002.
- Lerat (P.), 1994 : « Dérivation lexicale et dérive terminologique » dans *Méta*, XXXIX, 4, pp. 581-588.
- Lerat (P.), 1995 : *Les langues spécialisées*, PUF, « Linguistique nouvelle ».
- M'seddi (A.), 1984 : *qa:mu:s-al-lisa:nijja:(en arabe)*, Maison Arabe du Livre, dictionnaire de linguistique arabe français/français arabe, avec une introduction théorique sur la terminologie.
- Manuels scolaires : manuels de langue (7^e et 8^e de l'enseignement de base), ministère de l'Éducation nationale, CNP, Tunis (2000).
- Martinet (A.), 1966 : « Le mot », *Problèmes de langage*, Gallimard, collection « Diogène », pp. 39-53.
- Mehiri (A.), 1973 : *Les Théories grammaticales d'Ibn Jinnī*, Publications de l'Université de Tunis.
- Mejri (S.), Baccouche (T.), 1993 : *Du Mot*, Tunis, Dar Eljanoub (en arabe).
- Mejri, (S.), 2003 : « La traduction linguistique : problème terminologique ou construction conceptuelle ? » dans *Traduire la langue, Traduire la culture*, actes du colloque « Traduction humaine, Traduction automatique, Interprétation », Tunis 28-29-30 septembre 2000, pp. 177-190.
- Mounin (G.), 1993 : *Clefs pour la linguistique*, [trad. arabe Baccouche, (T)]. *mafa:ti:h-al-ʔalsunijja*, Éd. Sfidane. Tunis.
- Neveu (F.) et Ouerhani (B.), à paraître, « Les unités terminologiques de forme complexe en linguistique (étude comparée du français et de l'arabe) », Journées scientifiques tenues à Sousse en juin 2004, dans le cadre du projet AUF *Problèmes théoriques et épistémologiques de la traduction des terminologies linguistiques (français/arabe)*.
- Neveu (F.), 2004 : *Dictionnaire des sciences du langage*, Armand Colin, Paris.
- Ouerhani (B.), à paraître : « La problématique de l'exemple dans la traduction de la métalangue » dans Journées scientifiques *La Terminologie linguistique : problèmes traductionnels, conceptuels, épistémologiques*. Caen, 12 et 13 mai 2005.
- Saussure (de) (F.), 1985 : *Cours de linguistique générale*, [trad. Arabe Garmadi (S.), Chaouech (M.) et Ajina (M.)], Addaar Al-ḡarabya lil-Kitaab.

**TERMES ET TEXTES :
LA CONSTRUCTION DU SENS
DANS LA TERMINOLOGIE MÉDICALE**

Madalena Contente

departamento de educação médica, Faculdade de ciências médicas
Centro de Linguística da Universidade Nova de Lisboa
Universidade Nova de Lisboa
Lisboa, Portugal

Résumé : *Le texte scientifique dans le domaine de la médecine est un phénomène complexe où les unités terminologiques ont une fonction sémantique très importante. Le discours actualise le sens stabilisé des termes et permet la construction dynamique de nouvelles significations.*

Le corpus trilingue (portugais, anglais, français) est représentatif de différents types de discours et de niveaux de spécialisation. Les phénomènes étudiés sont la dénomination créative, la synonymie et l'équivalence interlinguistique. Ces phénomènes entretiennent souvent entre eux des relations d'interdépendance.

Mots-clés : *communication scientifique ; terminologie textuelle ; dénomination terminocréative ; synonymie différentielle ; équivalence interlinguistique ; traits sémantico-conceptuels*

1. INTRODUCTION

La description d'une langue de spécialité (Lsp) et, en particulier, de sa terminologie, dans une perspective de terminologie textuelle, est faite à partir de l'analyse de textes et de discours qui constituent des *corpus* spécialisés.

La communication scientifique et sa diffusion sont faites pour atteindre des objectifs, à partir de moyens et de structures institutionnelles de plus en plus diversifiées, ce qui intéresse les sociologues de la communication scientifique.

Notre recherche a comme objet la langue et la terminologie de la médecine. Nous avons, donc, constitué un *corpus* ouvert, selon les nouveaux critères soulignés par la linguistique de *corpus* ; en envisageant les différents types de discours ainsi que les différents moyens de divulgation. Il s'agit d'un *corpus* trilingue (portugais, français, anglais), qui intègre des textes scientifiques récents et représente différents types de discours et de niveaux de spécialisation. La constitution de ce *corpus* obéit à un ensemble de critères, à savoir : la dénomination créative ; la synonymie ; l'équivalence interlinguistique ; les utilisateurs prioritaires (étudiants et enseignants-chercheurs en médecine).

2. SOCIOLOGIE DE LA SCIENCE ET COMMUNICATION SCIENTIFIQUE

Aujourd'hui, selon les nouvelles perspectives de la sociologie de la science, la communication scientifique est sensible à la force illocutionnaire des propositions, aux stratégies de persuasion dans l'exposition des résultats, des négociations et des controverses, aux transformations de communication qui découlent du travail en équipe, de la vulgarisation scientifique et des différents niveaux de spécialisation. La communication scientifique est un processus de réception, interprétation et transformation de messages. Parfois, les réseaux scientifiques d'émetteurs fonctionnant en parallèle s'entrecroisent et les récepteurs, à leur tour, se transforment en nouveaux émetteurs, tout en reformulant les messages en fonction de leurs objectifs de communication. La complexité de la communication scientifique est inhérente au processus même de la science et aux différents acteurs qui participent à un domaine scientifique ou à une sphère d'activité.

Le texte scientifique dans le domaine de la médecine est un phénomène complexe où les unités terminologiques ont une fonction sémantique très importante. Le sens global des énoncés, donc des textes, résulte souvent du sens des unités terminologiques dans les énoncés, tout en rentrant dans un processus de construction du sens tout au long de l'énonciation. Cette construction du sens se fait par l'interaction de ce processus qui appréhende le niveau de l'énoncé. Les termes sont des outils stables de communication. Le discours actualise le sens stabilisé et permet la construction dynamique de nouvelles significations. Les différentes relations entre pensée et langage qui se produisent dans l'énonciation, débouchent dans une construction progressive du sens. Cependant, la construction progressive du sens dans l'énoncé relativise le rôle des catégories linguistiques et des catégories de la pensée. Par conséquent, les énoncés présentent des propriétés structurelles, liées à l'articulation entre densité sémantique et construction séquentielle, car les relations entre pensée et langage peuvent être formulées en fonction d'une projection et d'une conversion dimensionnelle de l'énoncé.

Les textes scientifiques présentent une thématique où les unités sémantiques dominantes sont les termes – éléments d'un système – qui actualisent leur signification dans les textes. Nous avons constaté, à travers l'analyse de notre *corpus* textuel, que les textes permettent une délimitation sémantique des termes ainsi que la création de nouveaux signifiants pour dénommer de nouveaux concepts.

La médecine, ainsi que toutes les autres sciences, entretient, aujourd'hui, des rapports étroits avec d'autres sciences ou des branches du savoir connexes. L'interdisciplinarité et la dynamique de la production scientifique internationale justifient, en grande partie, la diachronie rapide de la langue et de la terminologie médicales.

2.1. Représentation de la science

La représentation de la science (au sens de la sociologie des sciences), constitue une partie d'une image cohérente de l'activité scientifique d'une « idéologie de représentation » (Woolgar 1988) et d'un « idiome représentatif » (Pickering 1995) pour décrire une préoccupation métaphysique avec la correspondance entre la connaissance et leurs objectifs, comme l'affirment Lynche, M. et McNally, R. (1999 : 184).

Cette représentation concerne la transformation des objets en quelque chose de justifiable (cf. Lynche et McNally 1999 : 159-186). Les images et les descriptions sont des pratiques inhérentes à un domaine, mais elles ne représentent qu'une composante de l'ensemble.

J. Gayon (1998 : 69-70) se réfère à deux thèses épistémologiques en ce qui concerne le savoir scientifique. Dans la première, les énoncés indiquent ce qu'on doit faire pour obtenir un fait, dans la deuxième thèse, on justifie la science par son pouvoir de transformer le monde. Ces thèses instrumentalistes indiquent deux aspects de la science : les procédures de connaissance et leurs buts.

2.2. Dimensions de la communication scientifique

Aujourd'hui, la communication scientifique assume dans l'organisation de la recherche, sa dimension, dans l'optique d'une « nouvelle sociologie de la science ». On remarque la force illocutionnaire des propositions, en adaptant au travail scientifique la théorie des actes de parole (les propositions, inclusivement les propositions scientifiques, ne se limitent pas à leur valeur notionnelle et déclarative) ; on souligne l'importance des stratégies de persuasion dans l'exposition des résultats, des négociations et des controverses, des transformations de la communication du travail en équipe et de la divulgation scientifique. Du point de vue de la communication, on décrit la science comme un processus de réception, interprétation et transformation de messages. Une communication de plus en plus intense et complexe est inhérente au processus de la science.

3. LANGUE DE SPÉCIALITÉ ET COMMUNICATION

Selon Wimmer (1982 : 17, *apud* Balliu 2001 : 94), la langue de spécialité se différencie de la langue générale, à travers quelques aspects tels que la précision, l'univocité dénominate, l'économie et la relation matière/objet. Par contre, la langue générale est caractérisée par la polysémie, l'ambiguïté, la redondance et la multiplicité situationnelle et thématique.

Lothar Hoffmann (1984 : 53, *apud* Spillner 1992 : 44) affirme que la langue de spécialité est un ensemble de moyens linguistiques utilisés dans une situation de communication de spécialité spécifique afin d'assurer la communication entre leurs paires.

L'acquisition d'un savoir spécialisé dans le cadre d'une science donnée est possible à partir de textes qui permettent la communication et le transfert de connaissances spécialisées entre publics spécifiques et restreints. À ce sujet, Kocourek affirme que la « signification c'est communication virtuelle, communication c'est transmission de la connaissance signifiée », soulignant l'importance de l'interdisciplinarité entre spécialité et linguistique : « les textes spécialisés constituent la liaison interdisciplinaire permanente entre la spécialité et la linguistique. La langue de spécialité, textes et système, est, d'une part, l'instrument de la connaissance spécialisée – elle fait donc partie de la spécialité et devrait être citée ensemble avec les autres composantes de la spécialité ; elle est, d'autre part, le sujet de l'analyse linguistique » (Kocourek 1991 : 33).

L'analyse de la langue de spécialité, dans un contexte de communication de spécialité, doit tenir en compte les conditions pragmatiques de la situation de communication. Ainsi, au-

jour d'aujourd'hui, plusieurs investigations terminologiques se font dans une perspective d'une analyse de discours, écrit ou oral, de spécialité.

3.1. Terminologie et langue de spécialité

La terminologie est la composante la plus importante des langues de spécialité qui ont comme fonction le traitement et le transfert de connaissances spécialisées, c'est-à-dire qu'elles sont dépositaires d'une abondante terminologie. La terminologie est un ensemble cohérent de dénominations relatives aux concepts scientifiques et/ou techniques des langues de spécialité.

Selon E. Coseriu (1967 : 17) « on connaît les « signifiés » des terminologies dans la mesure où l'on connaît les sciences et les techniques auxquelles elles répondent et non pas dans la mesure où l'on connaît la langue ».

La langue spécialisée est un sous-système autonome qui a comme objectif la transmission de connaissances spécialisées, en situations de communication (écrites ou orales) relatives à des groupes socioprofessionnels. La langue de spécialité peut être observée et analysée dans les perspectives spécifiques des niveaux linguistiques : phonique, morphosyntaxique, syntaxique, lexical et textuel. La composante lexicale de la langue de spécialité comprend les unités lexicales spécialisées, les termes et les unités lexicales de la langue générale. Une langue spécialisée ne se réduit pas à une terminologie; les dénominations spécialisées (les termes) comprennent aussi des symboles non linguistiques.

La terminologie est un ensemble de termes, dénommant, dans une langue naturelle, les concepts d'un domaine de connaissance fortement thématisé.

3.2. Terminologie textuelle et processus discursif

La terminologie textuelle requiert que la recherche soit faite à partir de textes ou *corpus* spécialisés. Les *corpus* permettent une observation et une description linguistiques, de phénomènes terminologiques au sein d'un système linguistique.

La terminologie textuelle observe le terme en contexte, la polysémie, les variantes synonymiques, les néonymes, la syntaxe des textes spécialisés (cohésion, complexité, concision, condensation, impersonnalité) et les différents aspects sémantiques (cohérence, isotopie).

Cette nouvelle perspective permet une description adéquate des termes dans les plans : graphique, phonique, morphosyntaxique, morphosémantique, lexical, sémantique, discursif et pragmatique.

Le texte spécialisé (écrit ou oral) est la source et du travail en terminologie. Il constitue une source d'observation et de description des termes et dans les textes ; le texte est le dernier objectif car les termes créés sont utilisés en discours, en situations de communication spécialisée écrite ou orale.

La terminologie de la langue de spécialité constitue un sous-système de la langue, avec une dominante cognitive, où les textes ont pour finalité de signifier et communiquer, au sein d'une communauté restreinte de spécialistes; ces textes présentent des caractéristiques spécifiques soit graphiques soit syntaxiques et, surtout, un ensemble d'unités lexicales qui

acquièrent, dans les textes, une précision sémantique et métalinguistique: « la langue de spécialité est donc, dans ses textes, le trésor du savoir disponible et linguistiquement analysable, et, dans ses ressources, l'instrument du savoir futur » (Kocourek 1991: 42).

Ainsi, la méthodologie terminographique ne peut pas se restreindre aux nomenclatures des domaines scientifiques. La nomenclature permet d'énumérer, exhaustivement, un domaine. La logique et la cohérence du système sont plus importantes que la dénomination linguistique (par exemple, les unités symboliques de la nomenclature de la chimie). L'unité terminologique présente, parfois, des traits communs avec les noms du vocabulaire courant. Cependant, elle se caractérise par un nombre spécifique de traits sémantiques. Le phénomène de la dénomination est normalement articulé à la définition et à l'énonciation du discours spécialisé. La signification d'une dénomination ne peut être séparée du processus discursif de la situation de communication. Ainsi « partant du principe que les concepts scientifiques et techniques se matérialisent au moyen de formes linguistiques, les rapports terme/concept et les rapports entre les termes dans leurs cotextes sont envisagés sous un angle discursif, alliant le conceptuel au linguistique » (*cf.* Desmet 2002).

Le processus discursif actualise les termes en contexte, où se déroulent des facteurs de structuration textuelle qui contribuent à la cohésion discursive et textuelle. Des rapports entre les unités terminologiques et les différentes unités de liaison, résultent une cohérence discursive et textuelle et une structuration sémantique, qui relèvent de la fonction cognitive et communicative de la langue de spécialité.

3.3. Conceptualisation et terminogenèse

Les exigences d'une communication spécialisée en termes de concision et de cohérence requièrent le recours aux dénominations qui renvoient à la conceptualisation et au processus de terminologisation qui peut être observé dans la communication scientifique spécialisée à travers la création néologique (néonymie).

Les unités conceptuelles sont fondamentales dans un énoncé scientifique et le terme est la forme linguistique correspondant au concept qu'il nomme (*cf.* Kleiber 1984).

À l'intérieur d'un système conceptuel, les unités conceptuelles et, conséquemment, les unités terminologiques ou syntagmes terminologiques sont reliés par les relations interconceptuelles. En analysant ces relations inhérentes à l'énoncé scientifique, on constate deux espaces de représentation : l'espace linguistique constitué par des unités de langue, et un espace conceptuel dans lequel prédominent les unités conceptuelles, unités terminologiques et relations interconceptuelles. Ces unités établissent des relations dans le même système, elles requièrent une interliaison entre unités de langue et unités terminologiques, c'est-à-dire, entre le système linguistique et le système conceptuel. Cette relation interconceptuelle est directement liée aux aspects de nature sémantique, pragmatique et sociologique.

Dorothy Nakos (*cf.* 1995), faisant une analyse de l'objectivité des langues de spécialité, dans les textes didactiques du domaine de la médecine, réfère cinq caractéristiques : cohérence, précision, justification, neutralité et prudence. Cohérence dans la structure de l'organisation textuelle, précision de la terminologie, comme trait cohérent des langues de

spécialité, effacement du sujet de l'énonciation (médecin (s)), prudence marquée par l'objectivité. Ces aspects sont les éléments spécifiques des langues de spécialité, mais spécialement de l'objectivité textuelle de la science médicale. La motivation terminogénique de la médecine est directement liée à la motivation sémantique (métaphores, métonymies, etc.), par le biais d'éléments comme les morphèmes, les formants morphosyntaxiques et autres composantes des termes. Les différents types de formation terminogénique et la néonymie reflètent, en partie, les propriétés des termes qui existent dans la langue de spécialité.

Nous considérons le texte scientifique comme une organisation interconceptuelle où s'articulent le système de la langue et le système conceptuel, dans une circularité d'unités linguistiques et d'unités terminologiques et conceptuelles. Un texte, du point de vue cognitif, peut être envisagé comme un ensemble de connaissances et conceptualisations partagées par des intervenants dans la production et l'interprétation du texte et, simultanément, comme un processus d'introduction et de recueil de nouveaux éléments cognitifs (*cf.* Mateus *et alii* 1989 : 148). Ces auteurs (1989 : 146) avancent aussi qu'un des facteurs de textualité est la connexion conceptuelle qui résulte de l'interaction des éléments cognitifs des occurrences textuelles et de la connaissance que l'allocutaire a du monde ; dans ce cas précis, de son intellectualisation socioprofessionnelle.

4. DÉNOMINATION TERMINOCRÉATIVE ET MOTIVATION TERMINOGÉNIQUE

Nous avons constaté dans le *corpus* portugais une formation innovatrice des termes, ce que nous appelons « dénomination terminocréative », c'est le reflet d'une évolution constante et dynamique de la médecine, qui permet l'apparition de nouveaux termes construits, à notre avis, à partir d'une motivation terminogénique (*cf.* Contente 2003/4).

La plupart des unités lexicales des langues scientifiques sont motivées. Le signifiant des termes suggère, fréquemment, une partie de leur signification résultante de l'étroite articulation entre conceptualisation et langue scientifique. Les scientifiques créent les termes, justifiant, en général, les formes choisies : « la prédominance du motivé est si prononcée qu'elle est un caractère essentiel de la formation terminologique » (Guiraud 1978 : 98).

La motivation terminologique entretient une relation étroite avec la formation lexicale. La classification de la motivation, comparable à la classification de la formation lexicale, peut présenter quelques types : la motivation morphologique (dérivation, confixation, composition), la motivation syntagmatique (lexicalisation), la motivation par emprunt et par abréviation, la motivation sémantique (métonymie, métaphore, synecdoque, etc.) et la motivation phonomorphologique (Kocourek 1991 : 173).

La formation terminogénique constitue la base du terme simple et du terme complexe qui se présente, parfois comme un terme brachygraphique (Kocourek 1991 : 72-73). C'est le cas des exemples suivants :

[1] *Ca ++ intracelular*

[2] *CA 19.9/antígeno carcino-gastro-intestinal 19.9*

Les types de formation les plus fréquents sont la dérivation propre (nom + affixe), la confixation (confixe + confixe) et la composition (nom + nom). Dans cette typologie, on trouve des variantes et une formation plus diversifiée. De ces différents types de combinaisons de formation dérive un grand nombre de termes complexes et composés, comme c'est le cas des exemples présentés ci-dessous :

[3] *duodenopancreatectomia cefálica*

[4] *hipoparatiroidismo*

Cependant, dans un *corpus* multilingue on vérifie l'existence de phénomènes propres aux langues en contact : « deux langues sont en contact lorsqu'un locuteur (ou un ensemble de locuteurs) est amené à faire un usage simultané des deux dans des conditions données » (Pergnier 1989 : 23). Le terme « langues in contact » est employé en 1953 par Uriel Weinreich et, plus tard, développé par différents auteurs parmi lesquels : Maurice Pergnier, 1989 ; R. Kocourek, 1984, 1991 ; Isabel Rego, 1994.

Les cas de langues en contact résultent d'un processus psycholinguistique directement lié aux situations spécifiques de contact sociolinguistique – les interférences – donnant naissance à des énoncés résultant d'une influence d'un système linguistique sur un autre.

Les situations de communication spécialisée en médecine créent des conditions d'interférences entre systèmes linguistiques, desquelles résultent des calques phonologiques, morphologiques, syntaxiques, lexicaux et des emprunts. Prenons comme exemples les cas de figure suivants :

[5] *fracção de ejeção/ventricular ejection fraction*

[6] *síndrome do desaparecimento dos ductos biliares/vanishing bile ducts syndrome*

[7] *shunt bidireccional de Glenn*

La « terminocréativité » du portugais, dans la langue de spécialité de la médecine, résulte de fréquentes situations de contact avec les autres systèmes linguistiques, dans l'entourage de l'internationalisation de cette science, caractérisée par une circulation de connaissances, toujours, en actualisation. La dynamique de recherche n'est pas toujours accompagnée d'une création terminologique harmonisée, se révélant parfois un peu désordonnée. La divulgation d'un nouveau concept et de nouvelles découvertes et la compétition entre scientifiques ont comme conséquence la dynamique de cette langue de spécialité, en particulier au niveau de la terminologie, donnant naissance à de nouvelles formations en portugais au niveau terminosyntaxique, terminomorphologique et terminosémantique. C'est le cas des exemples qui suivent :

[8] *bypassar estenoses proximais da descendente anterior*

[9] *síndrome de Parkinson/síndrome parkinsonico/parkinsonismo/antiparkinsonico*

[10] *limpeza de resíduos esofágicos*

[11] *colectomia segmentar com lavagem anterógrada de Dubley*

La circulation du savoir scientifique, mise en œuvre par des revues de spécialité, par les nouvelles technologies de l'information, et maintenant, par l'Internet, permet une divulgation rapide des connaissances, qui contribue au développement d'une interlangue de spécialité.

5. SYNONYMIE DIFFÉRENTIELLE ET TRAITS SÉMANTICO-CONCEPTUELS

Un autre point étudié dans notre travail concerne la problématique de la synonymie dans la terminologie médicale. À ce propos, nous avons observé la synonymie et nous avons proposé l'analyse des termes synonymes à partir de ce que nous appelons la « synonymie différentielle » (cf. Contente 2003/4). Cette approche analytique requiert l'identification des traits sémantico-conceptuels qui spécifient la différenciation synonymique des termes.

Nous procédons à l'analyse descriptive, sémantique, conceptuelle et socioterminologique qui fournit les éléments fondamentaux pour établir la différence spécifique entre les différents synonymes.

Nous exemplifions la constitution de ces traits sémantico-conceptuels qui spécifient la différenciation synonymique des termes : < *trissomia 21* > < *síndrome de Down* > < *mongolismo* >.

<i>TRISSOMIA 21</i>	<i>SÍNDROME DE DOWN</i>	<i>MONGOLISMO</i>
maladie	maladie	maladie
malformation	malformation	malformation
congénital	congénital	congénital
chromosome 21	chromosome 21	chromosome 21
+ spécifique	- spécifique	courant
		désuet

Dans cet exemple, nous vérifions qu'il existe des traits qui sont communs et d'autres qui différencient les termes synonymes. Le terme < *trissomia 21* > est plus spécifique que les autres termes, car il nomme conceptuellement la cause de la maladie. Le terme < *síndrome de Down* >, est une dénomination éponymique qui présente un vide au niveau sémantico-conceptuel de la spécification de la maladie. Par conséquent il est moins précis dans la spécification sémantico-conceptuelle dénomminative. Le terme < *mongolismo* > porte un

trait familier de la dénomination de la maladie. Il ne s'agit pas d'une dénomination scientifique de cette langue de spécialité; il appartient à un niveau de non-spécialisation, car on l'emploie seulement quand le spécialiste communique avec un interlocuteur non spécialisé.

Nous procédons donc à une analyse dans une perspective de « synonymie différentielle », à travers l'identification des traits sémantico-conceptuels du terme-dénomination. Ainsi nous établissons les spécificités cognitives, inhérentes ou relationnelles du terme-concept.

Les traits sémantico-conceptuels sont représentés par des graphes qui sont subjacents à la conception d'un modèle de dictionnaire (cf. Contente 2003/4 b).

6. ÉQUIVALENCE INTERLINGUISTIQUE ET TRAITS SÉMANTICO-CONCEPTUELS

Pour établir une « équivalence interlinguistique » fiable, nous effectuons un exercice de terminologie comparée multilingue dans les trois langues de spécialité : portugais, anglais et français. L'identification de « l'équivalence interlinguistique » par le biais de l'identification des traits sémantico-conceptuels est axée sur les théories suivantes : la théorie du modèle relationnel de B. Pottier (cf. 1992) ; le schéma définitionnel de P. Thoiron, H. Béjoint et de Cl.-P. Boisson (cf. 1996) ; la sémantique différentielle de F. Rastier (cf. 1994).

En médecine, les traits sémantiques qui composent chaque terme correspondent aux traits conceptuels qui fournissent des éléments socioterminologiques (p. ex. terme désuet).

Nous délimitons les traits sémantico-conceptuels, en établissant les caractères ou traits distinctifs inhérents ou relationnels d'un terme, dans deux ou plusieurs langues, il s'agit d'une approche terminologique multilingue.

Prenons comme exemple le terme < *polineuropatia amiloidótica familiar* >. Il prend comme équivalent en anglais le terme < *familial amyloidotic polyneuropathy* > et comme équivalent en français < polyneuropathie amyloïde primitive >. Ces termes possèdent les traits sémantico-conceptuels suivants :

PORT.	ANG.	FR.
<i>Polineuropatia amiloidótica familiar</i>	<i>familial amyloidotic polyneuropathy</i>	polyneuropathie amyloïde primitive
polynévrite	polynévrite	polynévrite
amyloïde	amyloïde	amyloïde
transmission autosomique	transmission autosomique	transmission autosomique
pathologie portugaise	pathologie portugaise	pathologie portugaise

Le terme < *polineuropatia amiloidótica familiar* > a comme synonyme le sigle-terme < PAF >, sans équivalents dans les autres langues. Ainsi, nous établissons les traits pour ce terme sans avoir des équivalents dans les autres systèmes linguistiques :

	PORT.	ANG.	FR.
	PAF	< 0 >	<0>
	polynévrite		
	amyloïde		
	transmission autosomique		
	pathologie portugaise		
	sigle de « polineuropatia amiloidótica familiar »		

Le terme < *polineuropatia amiloidótica familiar* > a aussi comme synonyme < *doença dos pezinhos* > qui a comme équivalent en français le terme < mal de pesinhos >, sans équivalent en langue anglaise. Nous déterminons donc ainsi les traits pour ces termes :

	PORT.	ANG.	FR.
	<i>doença dos pezinhos</i>	< 0 >	mal de pesinhos
	polynévrite		polynévrite
	amyloïde		amyloïde
	transmission autosomique		transmission autosomique
	pathologie portugaise		pathologie portugaise
	courant		courant
			calque

Le terme < *polineuropatia amiloidótica familiar* > a encore comme synonyme < *doença de Corino* > sans équivalent en langue anglaise et ayant comme équivalent en français < maladie de Corino Andrade >. Les traits de ces termes sont les suivants :

PORT.	ANG.	FR.
<i>doença de Corino</i>	< 0 >	maladie de Corino Andrade
polynévrite		polynévrite
amyloïde		amyloïde
transmission autosomique		transmission autosomique
pathologie portugaise		pathologie portugaise
éponyme		éponyme
- spécifique		- spécifique
désuet		désuet

Les termes équivalents établis ont été analysés à partir des distinctions des différents traits sémantico-conceptuels de chaque terme dans les différents systèmes de langues. On a analysé aussi les synonymes intralinguistiques et interlinguistiques ainsi que les équivalents correspondants.

Nous avons établi ces « traits interlinguistiques » qui constituent des « univers d'expérience » conceptuels et cognitifs laissant des indices dans la langue (*cf.* Martin 1983 : 88). Il s'agit de représentations d'expérience, mais les « marqueurs linguistiques » peuvent avoir des formes différentes dans les langues naturelles (*cf.* Pottier 1992 : 71).

7. CONCLUSION

Les connaissances d'un domaine sont inscrites dans les textes produits par une communauté donnée en fonction de tel ou tel objectif de communication. À partir de l'approche textuelle des documents spécialisés, constituant notre *corpus* dans chaque langue, nous avons identifié ces phénomènes linguistiques avec la collaboration des chercheurs de la facultade de ciências médicas de l'Universidade Nova de Lisboa. Comme l'affirme Slodzian (*cf.* 2002), « c'est important le jugement de l'expert qui confronte ces unités à des connaissances déjà structurées, mémorisées et partagées par sa communauté ».

À partir de notre méthode, nous avons construit par la suite une ontologie qui correspond à l'étape de structuration des notions découlant de l'analyse contextuelle de notre *corpus*.

Enfin, nous avons construit un modèle de dictionnaire terminologique de médecine qui se fonde sur les connaissances conceptuelles et terminologiques. Ce dictionnaire est constitué par une base de données synonymiques et par une base de données relationnelles

d'équivalence interlinguistique, représentées par des graphes. Il s'agit d'un dictionnaire intralinguistique et interlinguistique qui présente des traits sémantico-conceptuels et des marqueurs linguistiques, en portugais, anglais et français (cf. Contente 2003/4 b).

BIBLIOGRAPHIE

- Bourigaut, (D.) et Slodzian, (M.), 1999 : « Pour une terminologie textuelle », dans *Terminologies nouvelles*, n° 19, Rint, pp. 29-33.
- Contente (M.) et Magalhães (J.), 2004 : « Typologie synonymique terminologique et synonymie différentielle », communication poster, Second Terminology Summit, UCLA Barcelona, nov. 2004, European Association for Terminology, actes sous presse.
- Contente (M.) et Magalhães (J.), 2005 : « L'équivalence interlinguistique dans la terminologie médicale », dans *actes du colloque De la Mesure dans les termes, hommage à Philippe Thoiron*, Université Lyon II, sous presse.
- Contente (M.) et Magalhães (J.), 2005 : « Sinonimologia e tipologia contrastiva da sinonímia terminológica em Medicina », dans *Debate Terminológico*, revista on-line, Riterm, União Latina.
- Contente (M.), 2003/4a : *Terminocriatividade, Sinonímia e Equivalência Interlinguística em Medicina*, dissertação de Doutoramento em Linguística, especialidade de Lexicologia, Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa, Abril de 2004.
- Contente (M.), 2003/4b : *Dicionário de Medicina – Sinónimos e Equivalentes*, Centro de Linguística da UNL, Faculdade de Ciências Sociais e Humanas, Departamento de Bioestatística e Informática – Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisboa.
- Coseriu (E.), 1967 : « Structures lexicales et enseignement du vocabulaire », dans *Les Théories linguistiques et leurs applications*, Strasbourg, Conseil de l'Europe, pp. 9-51.
- Depecker (L.), dir., 2005 : « La terminologie : nature et enjeux », dans *Langages*, n° 157, Paris, Larousse.
- Desmet, (I.), 2002 : « Synonymie et paronymie dans les langues spécialisées : théorie, méthodologie et applications », dans *Actes de la Journée Applications et implications en sciences du langage*, Université Paris III, le 16 mars 2002 (sous presse).
- Gayon (J.), 1998 : « Dagognet : quelle philosophe pour quelle histoire des sciences ? », dans Damien (R.), éd., *François Dagognet médecin épistémologue philosophe – une philosophie à l'œuvre*, Le Plessis-Robinson, Institut Synthélabo, pp. 59-78.
- Guiraud (P.), 1978 : *Les mots savants*, Paris, PUF, coll. « Que sais-je ? »
- Kleiber (G.), 1984 : « Dénominations et relations dénominatrices », dans *Langages*, n° 76 – La dénomination, Paris, Larousse, pp. 77-94.
- Kocourek (R.), 1991a : *La Langue française de la technique et de la science*, Wiesbaden, Oscar Brandstetter Verlag GmbH & Co. KG, 2^e édit.,
- Kocourek (R.), 1991b : « Textes et Termes », dans *Meta*, 36, n° 1, Montréal, Presses de l'Université de Montréal, pp. 71-75.
- Kocourek (R.), 1994 : « La terminologie et la linguistique textuelle », Porto, texte photocopié, Porto, 9-5-94.
- Lynche, (M.) et McNally, (R.), 1999 : « Aprisionando um monstro : a produção de representações num campo impuro », dans *A ciência tal qual se faz*, Lisboa, Edições Sá da Costa, pp. 159-186.
- Mateus, (M. H. M.), Brito (A. M.), Duarte (I.), Hub Faria (I.), 1989 : *Gramática da Língua Portuguesa*, Lisboa, Caminho, 2^e édit.,
- Nakos (D.), 1995 : « L'objectivité dans les langues de spécialité – cas particulier de la médecine de soins (étude comparée du français et de l'anglais) », dans *Meta*, vol. XL, n° 4, Montréal, Presses de l'Université de Montréal, pp. 701-706.
- Pergnier (M.), 1989 : *Les Anglicismes, danger ou enrichissement pour la langue française ?*, Paris, PUF.

- Pottier (B.), 1992 : *Sémantique générale*, Paris, PUF.
- Rastier (F.), 1994 : « Les sémantiques », dans Rastier (F.), *Sémantique pour l'analyse de la linguistique à l'informatique*, Paris, Masson, pp. 23-41.
- Rego Santos (I.), 1994 : *Línguas em contacto e estrangeirismos em vocabulários de especialidade*, Dissertação de Mestrado, Lisboa, FCSH, Universidade Nova de Lisboa.
- Slodzian, (M.), 2002 : « Terminologie et multilinguisme : des principes à l'application », dans Segong (F.) dir., *Multilinguisme et traitement de l'information*, Paris, Hermes Science-Lavoisier, pp. 89-109.
- Spillner (B.), 1992 : « Textes médicaux français et allemands – contribution à une comparaison interlinguale et interculturelle », dans *Langages*, n° 105, Paris, Larousse, pp. 42-65.
- Thoiron (Ph.) et Béjoint (H.), dir., 2000 : *Le Sens en terminologie*, Lyon, Presses universitaires de Lyon (Travaux du CRTT).
- Thoiron (Ph.), Béjoint, (H.) et Boisson (C.-P.), 1996 : « Notion d'« archi-concept » et dénomination », dans *Meta*, 41, n° 4, Montréal, Presses de l'Université de Montréal, pp. 512-524.

APPROCHE CONTRASTIVE DU *CORPUS* BILINGUE FRANÇAIS–ROUMAIN DU DOMAINE VÉTÉRINAIRE

Corina Georgeta Barna,
département relations internationales,
Université de sciences agricoles
et de médecine vétérinaire du Banat de Timișoara,
calea Aradului, n° 119, 300 645 Timișoara, Roumanie

Résumé : *L'objectif de cette étude est de relever des problèmes de la constitution et de l'exploitation du corpus bilingue français–roumain du domaine vétérinaire. Les méthodes utilisées correspondent à la terminologie textuelle, en utilisant l'analyse contrastive pour aboutir aux équivalences traductionnelles. Les résultats de la recherche sont représentés par un corpus terminographique de quelques centaines de pages et un corpus textuel de quelques milliers de pages. Les conclusions concernent les difficultés de l'exploitation du corpus représentatif.*

Mots-clés : *corpus, terminologie textuelle, vétérinaire, français, roumain*

1. CONTEXTE DES RECHERCHES

1.1. Enseigner le français langue étrangère aux étudiants des facultés de médecine vétérinaire

Pour l'enseignement du français langue étrangère (FLE) aux étudiants roumains du domaine vétérinaire, les besoins de ceux-ci concernent l'apprentissage de la terminologie à partir du macrocontexte, en employant des textes authentiques. Les documents choisis concernent les sous-domaines de l'anatomie, de la physiologie, de la zootechnie etc. et proviennent des cours de spécialité des facultés vétérinaires francophones, depuis plusieurs années. Le deuxième enjeu c'est qu'il n'y a pas un dictionnaire bilingue de spécialité. Ainsi, on fait appel à « la terminologie textuelle, une approche de construction de produits terminologiques (glossaires, terminologies, bases de connaissances terminologiques) qui repose sur l'analyse de *corpus* textuels » (Berland 2002 : 1).

Afin de constituer le *corpus* terminographique, on a utilisé pour le roumain *Mic dicționar de termeni medicali veterinari* (Taloș 1987) et *Dicționar de zootehnie și medicină veterinară* (Paștea 1969). Pour le français, on a constitué un *corpus* lexicographique informatisé, à partir des lexiques de spécialité du Web : *Dictionnaire du cheval*, *Dictionnaire vétérinaire du chien*, *Lexique pour les agents vétérinaires communautaires*, *Index alphabétique des noms des maladies parasitaires*, *Glossaire de reproduction et de sélection génétique*, *Glossaire de génétique moléculaire et génie génétique* etc. On a démarré un groupe de travail au niveau du traductème avec nos étudiants sur la base de données terminographique roumaine, pour établir les équivalences traductionnelles des unités lexicales françaises.

Un des problèmes de ce projet est de comparer les deux *corpora*. Il y a la solution de la numérisation du *corpus* roumain trouvé sur support papier, mais ce support papier n'est pas de meilleure qualité et le processus est très long à cause de la correction. On ne peut pas constater une vraie maniabilité de ce *corpus* terminographique roumain.

1.2. Traduire et interpréter pour le domaine vétérinaire

Les travaux de recherche ne concernent pas seulement le but didactique, mais aussi la cohérence traductive des scientifiques. Les chercheurs du domaine vétérinaire ont besoin d'un *corpus* textuel récent et pluridisciplinaire. La webographie leur offre des ressources, mais l'intégration des informations dans une base de données actualisée serait une meilleure solution. Le défi pour le linguiste du *corpus* est de leur ouvrir les portes sur un *corpus* textuel bilingue ou plurilingue adéquat et surtout informatisé.

Le sujet des recherches est représenté par la constitution et l'exploitation de ce *corpus*. Une terminologie multilingue offrirait aussi aux traducteurs et aux interprètes des ressources actualisées.

1.3. Implications des recherches

L'intérêt du développement d'une telle base de données concerne l'enseignement et la recherche, mais aussi l'intégration européenne et la sécurité alimentaire des produits du marché européen. Le langage du domaine vétérinaire contient aussi la terminologie de la sécurité alimentaire, ce qui nous concerne tous.

On peut également inclure parmi les intérêts les cooccurrences avec la médecine humaine, par exemple le sujet des zoonoses, qui regardent non seulement le bien-être des animaux, mais aussi la santé humaine. La médecine vétérinaire décrit au moins 7 espèces domestiques, ce qui la différencie de la médecine humaine.

Le premier niveau est didactique, les étudiants étant les premiers intéressés par ces recherches. Le deuxième niveau est scientifique, concernant les enseignants et les chercheurs du domaine vétérinaire, les deux niveaux constituant les implications dans le cadre institutionnel des facultés de médecine vétérinaire.

En ce qui concerne l'extension universitaire, il y a aussi des implications dans la société : les médecins vétérinaires de libre pratique ont besoin de ces connaissances pour leur formation continue. Toujours dans le cadre social, les consommateurs de services et d'aliments ont besoin des résultats des recherches.

Au niveau européen, les implications concernent la reconnaissance mutuelle des qualifications, la libre circulation des personnes et le marché des produits agroalimentaires. L'extension est évidente au plan mondial, car les aliments et les services de ce domaine circulent partout.

2. OBJECTIFS

L'objectif de cette étude est de relever les problèmes à résoudre par les spécialistes – linguistes et vétérinaires – pour la réalisation d'un *corpus* bilingue français – roumain du domaine vétérinaire, ayant également comme finalité une terminologie bilingue.

3. MÉTHODES

Pour l'exploitation des *corpora* – terminographique et textuel – le processus de documentation comprend aussi l'analyse comparative des *corpora* francophone et roumain. L'approche contrastive est nécessaire dans l'enseignement du français langue étrangère aux apprenants roumains, car les convergences des deux langues romanes mènent aux fautes interférentielles fréquentes (cf. Țenchea 1999 : 157). Il y a des fautes absolues, relatives et beaucoup de faux amis. Dans le langage médical, l'importance de la propriété des termes est vitale. L'approche contrastive est vraiment utile pour établir les équivalences traductionnelles.

4. RÉSULTATS

4.1. Constitution du *corpus* – hypothèses de travail

Dans le domaine didactique, on a constitué un *corpus* textuel français à partir du *curriculum* universitaire roumain. On a choisi ce *curriculum* parce qu'il a subi des transformations d'après le modèle français, au cours des négociations dans un programme phare qui s'est déroulé en 2002-2003, dans le cadre du jumelage institutionnel franco-roumain pour la reconnaissance mutuelle des qualifications à des fins professionnelles. À partir des diverses matières enseignées, on a constitué un *corpus* de milliers de pages, contenant la terminologie fondamentale. Le problème concerne le tri des textes, fait plutôt avec les critères de la linguistique et moins avec des critères du domaine concerné.

Dans le domaine scientifique, mes collègues vétérinaires ont signalé le besoin d'un *corpus* représentatif pour la terminologie, pour la documentation des étudiants et pour les spécialistes. Le *corpus* est constitué par les thèses de doctorat des écoles vétérinaires de France des dernières années. Ce *corpus* pourrait être considéré comme représentatif. Concernant le volume, les thèses de 2003 de l'École nationale vétérinaire de Toulouse dénombrent 161 titres et celles de 2004, 118 titres, les thèses de 2002 de l'École nationale vétérinaire de Lyon contiennent 219 titres, celles de 2003, 166 titres, les Écoles nationales vétérinaires de Maisons-Alfort et de Nantes ont aussi plus de 100 thèses de doctorat chaque année. Les thèses sont obtenues soit en téléchargeant directement depuis Internet en connaissant les URL, soit grâce aux relations de ma faculté avec les écoles vétérinaires de France. Deux bibliothèques nous renvoient chaque année les cédéroms avec les thèses de doctorat vétérinaire.

4.2. Propriétés du *corpus* textuel

On a analysé la représentativité du *corpus*, d'après les principes de signifiante, d'acceptabilité et d'exploitabilité (cf. Bommier – Pincemin 1999 : 415-419).

Concernant les conditions de signifiante, on considère ce *corpus* comme étant pertinent pour l'étude visée, car les textes choisis appartiennent à la littérature scientifique française du domaine vétérinaire, en abordant les thèmes les plus actuels du domaine. En ce qui concerne la cohérence, les textes sont homogènes, chacun d'eux développant une partie théorique, qui comprend la terminologie vétérinaire générale, et une partie contenant des applications pratiques, contenant des termes moins fréquents.

Les conditions d'acceptabilité concernent la représentativité. On peut considérer cet échantillon comme représentatif, car son langage a une nature redondante et on constate la clôture des unités textuelles, un trait caractéristique du langage médical. Néanmoins, Bidu-Vrănceanu (2002 : 28) signale la variation du contenu conceptuel en fonction des locuteurs dans la terminologie médicale et l'importance du contexte dans ce type de langage.

Un des traits de ce *corpus* est la régularité. Tous les textes ont été pris en considération. En ce qui concerne la complétude des textes, en choisissant une série de thèses de doctorat pour une année donnée, on peut considérer que le nombre de celles-ci couvre la plupart des matières enseignées ; si on fait une extension pour plusieurs années et pour plusieurs facultés, le degré de complétude est plus élevé.

Les conditions d'exploitabilité sont très importantes pour accomplir les objectifs d'analyse. Les textes sont homogènes : ils sont produits d'après un *Manuel du thésard*. Le volume des textes est essentiel pour l'enjeu de la représentativité du *corpus*. Si on a 100 titres pour une année, on aura plus de 2 000 titres pour 4 écoles pour cinq années, chaque thèse ayant environ 100 pages. Ainsi, on a plus de 200 000 pages.

La conclusion de cette analyse concernant la constitution du *corpus* est que le *corpus* obtenu a toutes les caractéristiques nécessaires pour être considéré comme un *corpus* spécialisé pour la terminologie française du domaine vétérinaire.

4.3. Exploitation du *corpus*

« Le *corpus* est indissociable de l'analyse » (Dalbera 2002 : 9).

Le problème qui se pose concerne la propriété intellectuelle. L'exploitation des documents par leur introduction dans un *corpus* informationnel nécessite l'autorisation de reproduction. Pour constituer seulement une terminologie ou un dictionnaire bilingue à partir du *corpus* textuel il n'y a aucun problème, mais pour offrir des macrocontextes il y aura des problèmes légaux. La constitution d'une base de données avec les macrocontextes doit se faire à partir des documents, avec l'accord des auteurs.

La solution est de réaliser une communauté de discours, où « le contenu peut être justifié comme étant « représentatif » d'une communauté définissable » (cf. Williams 1999 : Résumé). La solution est de faire des *corpora* informatisés dans des projets interinstitutionnels, dans des projets internationaux pour chaque domaine concerné. On pourrait réaliser ce projet dans la communauté scientifique, pour pouvoir résoudre les problèmes révélés par les recherches pour plusieurs langues.

5. CONCLUSIONS

5.1. Implications des recherches

Dans le contexte du développement de l'espace francophone scientifique et pour la mise en œuvre de la déclaration de Bologne, l'importance de l'enseignement du français langue étrangère dans la formation des étudiants va croissant. Par ailleurs, la terminologie vétérinaire recoupe la terminologie de la sécurité alimentaire, ce qui a des applications pratiques dans la sécurité des produits alimentaires du marché européen. Mentionnons également les zoonoses, les maladies communes aux humains et aux animaux.

L'enseignement du français langue étrangère pour les étudiants du domaine vétérinaire de Roumanie représente beaucoup d'enjeux. Cela implique des problèmes concernant la constitution du *corpus* spécialisé pour l'étude systématique de la terminologie. Les solutions se trouvent dans la collaboration de l'enseignant du FLE avec le corps enseignant roumain et francophone ainsi qu'avec les étudiants. Mais la plus grande ouverture est représentée par le développement du Web, plus précisément l'existence des ressources textuelles virtuelles et la possibilité offerte aux étudiants et aux professionnels d'étudier en ligne la dynamique du langage scientifique.

Du point de vue de la didactique des langues, la méthode d'analyse contrastive des *corpora* nouvellement constitués offre, à côté d'autres méthodes, des ouvertures pour la création des ressources terminologiques dynamiques, en fonction des nouveaux contextes virtuels. La constitution d'un *corpus* vétérinaire bilingue français – roumain pour l'étude contrastive de la terminologie présente de nombreuses particularités liées à la didactique.

5.2. Enjeux et solutions

Le problème de l'exploitation des *corpora* écrits, c'est que le travail de numérisation les textes prend beaucoup de temps. La solution serait de former dès le début des *corpora* informatisés et de choisir les meilleures solutions pour la normalisation et pour l'exploitation. Le sujet de ces recherches est d'identifier les aspects de la constitution, la normalisation et l'exploitation de ces *corpora*.

Les scientifiques, les étudiants et les professionnels ont besoin d'un *corpus* textuel récent et pluridisciplinaire pour trouver une cohérence traductive. Construire un *corpus* représentatif pour réaliser une terminologie en macrocontexte est réalisable, mais publier des documents authentiques implique des problèmes concernant la propriété intellectuelle. La solution est de constituer un *corpus* informatisé bilingue ou multilingue, en demandant aux scientifiques du domaine de produire ou de fournir des textes représentatifs, d'après un *curriculum* fourni par eux ; il en résultera une communauté de discours scientifique, qui va réaliser un *corpus* spécialisé.

Ce *corpus* de référence, un ensemble documentaire structuré, pourra répondre à des objectifs variés. On pourrait également étudier des particularités du discours scientifique médical vétérinaire, en fonction du type de discours, si on ajoute aussi d'autres types de texte. Les résultats seraient utiles aux professionnels.

Un travail dans une communauté de chercheurs, dans un projet pour la constitution d'un *corpus* informatisé bilingue ou plurilingue du domaine vétérinaire serait une solution qui pourrait servir aux traducteurs, aux interprètes, aux lexicologues, aux lexicographes, aux enseignants de langues étrangères, aux scientifiques et, évidemment, aux étudiants et aux professionnels du domaine. Les entreprises pourraient également fournir des ressources et exprimer leurs besoins ou leurs avis, et soutenir ce réseau qui constituerait une ressource pour la formation continue de leurs employés.

BIBLIOGRAPHIE

- ***, 2003 : De la protection à l'exploitation d'un ensemble informationnel, dans *JurisTIC* [en ligne], disponible sur : <http://www.juristic.net/article73.html>
- ***, 2003 : Élaboration d'un *corpus* informationnel – règles et devoirs, dans *JurisTIC* [en ligne], disponible sur : <http://www.juristic.net/article78.html>
- ***, 2003 : *Thèses de doctorat vétérinaire Toulouse 2003*, École nationale vétérinaire de Toulouse, [cédérom].
- ***, 2004 : *Thèses de doctorat vétérinaire Toulouse 2004*, École nationale vétérinaire de Toulouse, [cédérom].
- ***, 2005 : Fonds documentaire, École nationale vétérinaire de Nantes, [en ligne], disponible sur : <http://wwwbibli.vet-nantes.fr/Fondsdoc.htm>
- ***, 2005 : *Recherches des thèses d'exercices soutenues dans les quatre ENV*, École nationale vétérinaire de Toulouse, [en ligne], disponible sur : <http://www.envt.fr/Documentation/page621.php>
- ***, 2005 : *Texto !* [en ligne], disponible sur : http://www.revue-texto.net/T_index.html
- ***, 2005 : *Thèseindex*, École nationale vétérinaire d'Alfort, [en ligne], disponible sur : <http://www.vet-alfort.fr/bibliotheque/bib1/thseinde.htm>
- ***, 2005 : *Thèses de Lyon soutenues*, École nationale vétérinaire de Lyon, [en ligne], disponible sur : <http://www.vet-lyon.fr/bib/fondoc/theselyon.htm>
- Barna (C.), 2003 : « Internet dans la classe de Français langue étrangère pour les étudiants vétérinaires », dans *Scientific Papers. Veterinary Medicine*, vol. XXXVI, Timișoara : Universitatea de Științe Agricole și Medicină Veterinară a Banatului din Timișoara (USAMVBT), pp. 666-673.
- Barna (C.), 2003 : « Sources de documentation via Internet en zootechnie et en médecine vétérinaire », dans *Scientific Papers. Animal Sciences and Biotechnologies*, vol. XXXVI, Timișoara : USAMVBT, pp. 573-578.
- Barna (C.), 2003 : « Termes spécifiques pour la parturition aux animaux », dans *Scientific Papers. Veterinary Medicine*, vol. XXXVI, Timișoara : USAMVBT, pp. 674-677.
- Barna (C.), 2004 : « Analyse contrastive des dénominations roumaines et françaises des animaux domestiques », dans *Scientific Papers. Veterinary Medicine*, vol. XXXVII, Timișoara : USAMVBT, pp. 954-958.
- Barna (C.), 2004 : « L'espèce et le genre des animaux domestiques en latin, roumain et français – Approche étymologique », dans *Scientific Papers. Veterinary Medicine*, vol. XXXVII, Timișoara : USAMVBT, pp. 948-953.
- Berland (S.) et Grabar (N.), 2002 : *Assistance automatique pour l'homogénéisation d'un corpus Web de spécialité*, JADT 2002 : 6^{es} Journées internationales d'Analyse statistique des données textuelles.
- Bidu-Vrănceanu (A.), 2002 : *Lexic comun, lexic specializat*, Universitatea din București, [en ligne], disponible sur : <http://www.unibuc.ro/eBooks/vranceanu/index.htm>
- Bidu-Vrănceanu (A.), 2002 : *Principii generale privind corpusul lexicului științific interdisciplinar*, Universitatea din București, [en ligne], disponible sur : <http://www.unibuc.ro/eBooks/filologie/lexic/capa.htm>

- Bied-Charreton (N.), 2002 : *Thèses 2002*, Bibliothèque ENVT, École nationale vétérinaire de Toulouse, [cédérom].
- Bommier-Pincemin (B.), 1999 : *Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents*, thèse de doctorat en linguistique, Université Paris IV-Sorbonne, 6 avril 1999, chapitre VII : « Caractérisation d'un texte dans un corpus : du quantitatif vers le qualitatif », § A « Définir un corpus », pp. 415-427.
- Chiș (D.), 2002 : « Terminologie et constitution disciplinaire », dans Superceanu (R.), Lungu-Badea (G.), Dejica (D.) et Petzek (M.), édit., *Comunicare profesională și traductologie 2002, Lucrările conferinței internaționale 26-27 septembrie 2002 Timișoara, România*, Timișoara : Éd. Orizonturi Universitare, pp. 223-228.
- Constantinovi (S.), 2002 : « Terminologia : între lexicologie și lingvistică aplicată », dans Superceanu (R.), Lungu-Badea (G.), Dejica (D.) et Petzek (M.), édit., *Comunicare profesională și traductologie 2002, Lucrările conferinței internaționale 26-27 septembrie 2002 Timișoara, România*, Timișoara : Éd. Orizonturi Universitare, pp. 113-119.
- Dalbera (J.-Ph.), 2002 : « Le corpus entre données, analyse et théorie », dans *Corpus*, n° 1 : « Corpus et recherches linguistiques », novembre 2002.
- Deroubaix (J.-C.), 2004 : *Que faire des corpus multilingues parallèles ? Une expérience*, JADT 2004 : 7^{es} Journées internationales d'analyse statistique des données textuelles, pp. 295-303.
- Drouin (P.), 2004 : *Spécificités lexicales et acquisition de la terminologie*, JADT 2004 : 7^{es} Journées internationales d'analyse statistique des données textuelles, pp. 345-352.
- Frey (C.), 1997 : « Corpus et information », dans Frey (C.) et Latin (Danièle), édit., *Le Corpus lexicographique. Méthodes de constitution et de gestion*, actes des troisièmes journées scientifiques du réseau thématique de recherche « Étude du français en Francophonie », AUPELF-UREF, Louvain-la-Neuve : De Boeck-Larcier.
- Gheorghiu (M.), 1981 : *Probleme de tipologie contrastivă a limbilor. Determinanți congruenți de relație*, București : Ed. Științifică și Enciclopedică, pp. 7-36.
- Habert (B.), 1999 : *Un corpus clef pour le français actuel*, Présentation du projet Corpus CLEF, [en ligne], disponible sur : <http://www.biomath.jussieu.fr/CLEF>
- Ioani (M.), 2002 : « Modélisation dans le français de la communication scientifique et technique », dans Superceanu (R.), Lungu-Badea (G.), Dejica (D.) et Petzek (M.), édit., *Comunicare profesională și traductologie 2002, Lucrările conferinței internaționale 26-27 septembrie 2002 Timișoara, România*, Timișoara : Éd. Orizonturi Universitare, pp. 229-236.
- Jaubert (A.), 2002 : « Corpus et Champs disciplinaires. Le Rôle du point de vue », dans *Corpus*, n° 1 : « Corpus et recherches linguistiques », novembre 2002.
- Labatut (P.) et Bied-Charreton (N.), 2001 : *thèses 2001*, Bibliothèque ENVT, Ecole nationale vétérinaire de Toulouse, [cédérom], version 1.0.
- Mellet (S.), 2002 : « Corpus et recherches linguistiques », dans *Corpus*, n° 1 : « Corpus et recherches linguistiques », novembre 2002.
- Paștea (E.), Dinu (I.), Bârză (H.) et Berehoiu (G.), 1969 : *Dicționar de zootehnie și medicină veterinară*, București : Éd. Agrosilvică.
- Petrescu (C.), 2002 : « A contrastive approach to translation », dans Superceanu (R.), Lungu-Badea (G.), Dejica (D.) et Petzek (M.), édit., *Comunicare profesională și traductologie 2002, Lucrările conferinței internaționale 26-27 septembrie 2002 Timișoara, România*, Timișoara : Éd. Orizonturi Universitare, pp. 316-320.
- Pincemin (B.), 2002 : Compte rendu du n° 2 de la revue *Corpus* sur « La distance intertextuelle », *Texto !* Décembre 2004 [en ligne], disponible sur : http://www.revue-texto.net/Parutions/CR/Pincemin_CR.html
- Sautermeister (C.), 1989 : « Pour une meilleure compétence lexicale », dans Ibrahim (A. H.) et Zalessy (M.), coord., *Lexiques*, Rennes, Hachette, pp. 122-133.
- Taloș (V.), 1987 : *Mic dicționar de termeni medicali veterinari*, Timișoara : Institutul Agronomic Timișoara — secția Medicină Veterinară, Societatea de Medicină Veterinară – Filiala Timiș.

Țenchea (M.), 1999 : *Études contrastives (français-roumain)*, Timișoara : Éd. Hestia.

Williams (G.), 1999 : « Les réseaux collocationnels dans la construction et l'exploitation d'un *corpus* dans le cadre d'une communauté de discours scientifique », dans les *Annuaire ATALA* – Association pour le traitement automatique des langues [en ligne], disponible sur : http://www.atala.org/article.php3?id_article=209

LE LANGAGE DE LA MÉDECINE : LES MOTS POUR LE DIRE

Christian Balliu

ISTI

Haute École de Bruxelles

Bruxelles, Belgique

Résumé : *La médecine, comme tout domaine de spécialité, se caractérise par une terminologie particulière. Celle-ci pourrait donner à penser que le langage médical est dénoté, strictement scientifique, afin de garantir la rigueur et l'exactitude des faits exposés. En réalité, le discours médical est tout aussi professionnel que savant ; il recourt ainsi à des figures métonymiques et métaphoriques qui connotent le discours et donnent vie aux termes en véhiculant une véritable philosophie de l'art de guérir.*

Mots-clés : *médecine, notion, sociolecte, connotation, subjectivité*

On ne trouvera, dans cet exposé, aucune question de théorie, aucun débat de linguistique ou de terminologie, mais seulement une étude de cas et la recherche d'un mobile.

Avec le dessein de ne m'adresser qu'à des linguistes et à des étudiants, non à des praticiens de l'art de guérir, je ne pouvais pas croire que, même dans le cadre d'une conférence, un ton et des démonstrations trop arbitrairement imperturbables dussent être adoptés. Partout où la nécessité m'en est apparue, j'ai délaissé le souci architectural de l'exposé, la « rotondité » de la question, et préféré montrer, non pas en des édifices didactiques superficiels, mais dans leur état brut, et non encore encadrés, les faits eux-mêmes.

Enfin, sans ignorer ce que l'absence réelle ou apparente de recherche bibliographique, de recours aux travaux des autres, donne de mouvement au récit et de facilités à l'auteur et au lecteur, je n'ai pas songé à me priver ni à priver celui qui me lira de l'instruction éparse dans les meilleurs travaux sur le sujet. Usant d'eux de mon mieux, je ne savais plus ne pas citer les auteurs. Bref, je ne ferai pas comme Voltaire, qui conseillait de « citer tout le monde, sauf ceux à qui on a pris quelque chose » ; je paierai volontiers mes dettes intellectuelles.

La médecine est une discipline qui a son propre langage. Cela est trivial. Cette corrélation entre discipline et langage n'est évidemment pas sans conséquence sur l'évolution de celle-ci, de ses notions et des dénominations chargées de relayer ces notions au sein de la société. Elle conduit à mon sens à trois paradoxes fondamentaux.

Derrière les notions et la terminologie du domaine se cachent des choix théoriques dont l'intuition n'est pas absente. Le grand Claude Bernard s'était un jour adressé à son élève Paul Bert de la manière suivante : « Laissez votre imagination avec votre paletot au vestiaire, mais reprenez-la en sortant ». J'en arrive ainsi tout de suite à mon premier paradoxe :

l'objectivité qui est à la base de toute science doit nécessairement s'adosser en l'espèce à une subjectivité inébranlable qui est la caution de l'intuition, de la découverte, du génie. Comment une science qui s'occupe des arcanes physiologiques, mais aussi psychologiques, d'un sujet pourrait-elle ne pas être, en partie du moins, marquée au sceau de la subjectivité ? Comment un médecin pourrait-il investiguer le corps d'un patient en faisant abstraction de ses propres grilles de lecture ?

Les mots traduisent cet état d'esprit. Ils donnent aux termes leur sens en les emplissant de vie. En traduction médicale, être à l'affût des sentiments, des sensations, c'est être à l'écoute du sens. Si les termes font appel inconsciemment à la langue répertoire, les mots organisent l'univers physiologique selon des critères holistiques. La question fondamentale est la suivante : le discours médical cautionne-t-il le caractère scientifique de la discipline par une terminologie, à savoir un lien indéfectible entre une notion et une dénomination, ou ce discours fait-il aussi la part belle aux mots, qui mettent en jeu un lien elliptique, métonymique, voire métaphorique, entre la notion et sa matérialisation linguistique ?

Ayant observé de nombreux dialogues dans des services hospitaliers de pointe, ayant parcouru de nombreux protocoles et diagnostics transmis de confrère à confrère, *a fortiori* dans un même domaine de spécialité, j'ai pu remarquer le caractère central du langage naturel spontané dans les processus de communication entre médecins. Il y a une oralité dans les communications écrites entre spécialistes. Le langage médical est pour nombre de médecins un « parlécrit » pour reprendre l'expression de Jeay (1991 : 47).

En n'ancrant pas la médecine dans une sociologie diachronique par un souci d'actualisation, la terminologie a souvent préféré dans ses analyses des domaines de spécialité le linguistique au détriment du discursif. C'est cette approche méthodologique, fondée sur une classification en tiroirs, qui a été privilégiée par la terminologie pour éluder les vases communicants polysémiques. Il s'agit dans son esprit de circonscrire l'usage d'un terme ou d'un syntagme à un domaine particulier dont il ne pourrait s'échapper en gardant le même sens. Le sens équivaldrait dès lors à une signification prise dans l'état d'un domaine de spécialité au monologue singulier. Or, la monosémie d'apparence cache quelquefois des sens plus profonds, imperceptibles si on ne lit pas sous la surface des signes linguistiques qui sont, eux aussi, la simple matérialisation d'enjeux sémantiques et sémiologiques plus enfouis. J'en arrive alors à mon deuxième paradoxe. Si l'on devait accrédi-ter la thèse d'une terminologie scientifique, monoréférentielle et univoque, la distinction entre approches synchronique et diachronique n'aurait aucun sens. Les termes seraient les garants intemporels d'un lien notionnel qui traverse les époques sans dommages, dans la mesure où la constante dénominative serait insensible aux outrages du temps. Ce n'est assurément pas le cas.

En d'autres mots, la terminologie médicale dépend étroitement de l'environnement social qui la voit naître. Elle relève des modes de raisonnement qui induisent les évolutions techniques plutôt qu'ils ne les suivent. L'hyperspécialisation actuelle fait que même l'interniste peut être désarmé devant un patient d'une cinquantaine d'années qui lui commente un résultat ou lui cite un nom propre dont il n'a jamais entendu parler. Être spécialisé, c'est aussi méconnaître des pans entiers du savoir médical. C'est là toute la difficulté de rédiger des

dictionnaires et des bases de données terminologiques qui conjuguent harmonieusement actualisation du domaine et vision épistémologique large.

Cela ouvre la voie au sociolecte médical, lequel véhicule aussi une théorie du langage des fonctions qui résiste aux tentatives de hiérarchisation et d'inclusion. Le vocabulaire et les modes d'expression de la médecine sont *professionnels*, c'est-à-dire qu'ils ne sont pas nécessairement *savants*. Cela explique également que ce vocabulaire se renouvelle sans cesse ou, mieux, qu'il est renouvelé par la communauté scientifique au fil des découvertes ou des modes. C'est ainsi qu'avec le temps la phtisie s'est muée en tuberculose et que l'anaphylaxie est devenue l'allergie. C'est la compréhension intime du *modus operandi* de l'affection qui en change la dénomination. Mais l'évolution médicale n'induit pas nécessairement une correction lexicale. L'hystérie n'est bien entendu plus considérée comme à l'époque de Charcot (1889) et son étiologie « utérine » si je puis dire est depuis longtemps délaissée. Le terme n'a pourtant pas changé.

Même au plan synchronique, le discours médical se caractérise – et c'est le cas de nombreuses activités humaines – par une *compréhension opérative* comme l'écrit Ochanine (1978 : 63-79), à savoir la mobilisation cognitive d'un expert dans une situation qui lui est habituelle, dans un domaine circonscrit. Au plan psychologique, une grille de lecture propre au domaine envisagé est déjà activée par l'interlocuteur avant même que l'information ne lui soit fournie. Le langage opératif, ce dialecte particulier aux experts, est économique, il évite la redondance en faisant l'impasse dans le discours sur les connaissances partagées. L'économie cognitive autorise le « devenir » (j'insiste sur cette notion) monosémique du terme.

Cette compréhension opérative a été décrite par Chibout, Kotowicz et Briffault (2003 : 103) :

- le *laconisme* et l'*adéquation à la tâche* : seuls les éléments pertinents de l'objet par rapport à la tâche en cours sont pris en compte ;
- la *déformation fonctionnelle* : certains traits de l'objet, de la situation ou de l'opération sont accentués par rapport aux autres.

Les outils par excellence de ce discours sont les raccourcis de langage d'une part, et la figure métonymique ou elliptique d'autre part, le procédé métaphorique enfin.

Voici l'un ou l'autre exemple de raccourci :

- les acronymes : ASP pour « abdomen sans préparation » (en radiologie) ; UIV pour « urographie intraveineuse » ;
- des abréviations, souvent des apocopes : « nucléo » pour « nucléolyse » ; « endo » pour « endoscopie » ; « bru » (voire « rub ») pour « René Brugière », le patient qui permit à Luc Montagnier d'identifier le VIH à l'institut Pasteur de Paris en 1983.

La figure métonymique ou elliptique :

- « Il y a encore un foie et un rein qui vont monter » (en salle d'opération) ;
- « Le transit est sur la table » (en radiographie) ;

– « Amenez l’ulcère pour un contrôle ».

Il est à noter que dans ces exemples la métonymie se rapporte chaque fois au patient. Mais ce n’est pas toujours le cas, comme le montre le phénomène de polysémie suivant dans le cadre d’un service de radiologie :

– « L’infirmière est allée chercher le foie du quatrième » (patient) ;

– « Affichez-moi le foie » (cliché sur le négatoscope) ;

– « Il y a un foie qui vient de tomber » (demande de radiographie) ;

– « Il faut retaper ce foie » (le protocole d’examen clinique contient des coquilles).

On peut dire que nous sommes en présence à la fois de contractions du signifiant par l’utilisation d’acronymes et d’abréviations, et de contractions du signifié par le jeu des métonymies.

Cette notion linguistique de contraction du signifiant et du signifié a bien entendu un relais dans la pensée du praticien : la dépersonnalisation du patient et, en aval, son appropriation par le corps médical (Balliu 2002 : 77-101).

Les exemples cités le montrent bien : l’ulcère, le foie, le rein... Une théorie du langage des organes s’est substituée à une vision holistique de l’homme. C’est à la fois la rançon d’une médecine hyperspécialisée, divisée en sous-domaines, et qui ventile par la même occasion l’homme en territoires anatomiques de spécialité. Même lorsque le thérapeute s’adresse directement au patient, il le subordonne à une vision organique de l’art de guérir. Un exemple à mon sens extraordinaire : « Surveillez votre foie ! ». L’organe se voit doté ici d’une personnalité morbide. Parfois aussi, le patient se voit dénué de toute existence autonome comme dans : « Amenez le 212 en salle d’op’ ».

La terminologie médicale a été hantée ces dernières années par le fantôme de l’émiettement. Le foisonnement des notions, la multiplication des dénominations et la complexité croissante des connaissances ont poussé les terminologues à subdiviser les domaines en sous-domaines dans un premier temps, en microdomaines ensuite. Le savoir médical devient extérieur au sujet, et le médecin en serait le seul dépositaire en vertu d’une compétence intrinsèque. C’est le thème de l’appropriation de la maladie et du patient par le médecin. Par exemple :

« *Un malade de Launay* est pris brutalement, trois heures après le repas, d’une douleur extrêmement violente, qui se calme pendant tout le lendemain ; mais brusquement, dans la nuit, la douleur réapparaît. À l’opération, on trouve une perforation cachée sous des fausses membranes. Le second temps peut être tardif, comme dans *un cas de Courty*, où l’on voit, après un syndrome de perforation, un abcès prégastrique évoluer, guérir, par l’incision, et la péritonite ne survenir que dans une deuxième crise, plusieurs mois après » (Mondor 1949 : 225).

Un autre cas d’appropriation du patient par le médecin est l’éponymie : la maladie d’Alzheimer, le bacille de Koch, le sarcome de Kaposi, la maladie de Parkinson... Tous des médecins qui volent aux patients la « paternité » de leur pathologie et qui, dans un

nombre non négligeable de cas, ne peuvent pas même revendiquer la paternité de leur guérison.

La communication médicale vise indubitablement des créneaux, des crédits de recherche, des débouchés ; elle est autant orientée vers un positionnement dans la communauté scientifique que par la seule volonté de faire progresser la discipline. La soif de notoriété, le statut universitaire, la reconnaissance des pairs, autant d'alibis qui poussent le spécialiste à se revendiquer de son propre mouvement, à chercher tantôt un père spirituel, tantôt des disciples, afin de créer une école de pensée ou du moins y appartenir. La médecine, c'est aussi cela : une lutte d'influences et de pouvoir. À cet égard, la terminologie joue un rôle capital, car elle symbolise et labellise la notion dans la société (Balliu 2005 : 1-19).

C'est plutôt le mot, voire le nom dans le cas de l'éponymie, associé à une découverte ou à une avancée qui marque les esprits que le progrès en lui-même. Voici un exemple que je tire du merveilleux *Diagnostics urgents-Abdomen* d'Henri Mondor, qui fut pendant de longues années professeur de clinique chirurgicale à la Salpêtrière. À propos de la douleur caractéristique des ulcères perforés, il écrivit ceci :

« La douleur abdominale, soudaine, atroce, angoissante, *le coup de poignard péritonéal de Dieulafoy*, est le premier symptôme de la perforation » (1949 : 194).

On remarquera le cousinage intellectuel fascinant de la terminologie éponymique et de sa définition à l'aide de simples mots.

Le summum de l'appropriation se retrouve dans les génitifs d'appartenance, souvent dérivés des génitifs saxons, comme dans la *listériose* (*Lister's disease*), du nom du médecin anglais Joseph Lister, professeur à King's College à Londres. Rien de scientifique dans tout cela et une paternité parfois toute relative, dans le sens où l'éponyme peut changer selon le pays.

Chose curieuse, le nom du découvreur de la maladie (ou de son guérisseur) peut en effet varier selon les pays et la maladie, la même, porter des noms différents selon les latitudes. C'est ainsi que le goitre exophtalmique hyperthyroïdien se dénomme *Graves' disease* dans les pays anglo-saxons, en l'honneur de l'interniste irlandais Robert James Graves, alors qu'en Allemagne cette même maladie s'appelle *Basedow Krankheit*, d'après le médecin allemand Karl Adolf von Basedow. En Italie, la même maladie est dénommée *malattia di Flajani*, en hommage au médecin italien Giuseppe Flajani. En France, pays resté en marge de cette dispute d'appartenance, l'affection porte généralement le nom de *maladie de Basedow-Graves* (Navarro 1997 : 10).

Ce phénomène n'est pas linguistique. Sa matérialisation dans le discours est symptomatique du vécu quotidien du praticien, lequel doit affronter la méfiance et l'opposition de ses confrères, lorsque lui-même ne les leur fait pas subir. Plus la notoriété est en jeu, plus l'attitude scientifique semble devoir s'effacer au profit de la position dans la société. Pour le dire autrement, l'effet d'annonce l'emporte sur l'annonce des faits. À titre d'anecdote, je raconterai l'effervescence médicale qui entoura la mort de Léon Gambetta (décédé d'appendicite rétro-cæcale non diagnostiquée) et qui n'est pas sans rappeler les savoureuses mises en scène de la Faculté par Molière. Autour de l'illustre patient se trouvaient de non

moins célèbres praticiens : Charcot, Siredey, Verneuil et Lannelongue. L'épisode se passe en 1882 et voici ce que consigne Lannelongue dans ses *Cliniques chirurgicales* de 1905 :

« Une opération aurait-elle sauvé le malade ? je ne saurais le dire, mais on devait et il fallait la pratiquer. Ma conviction, là-dessus, était si profonde et si absolue que je tentai, en dehors du moment de nos consultations, plusieurs démarches auprès de Charcot, de Verneuil, de Trélat pour les y déterminer. Je croyais l'opération si nécessaire et si urgente que j'avais porté sur le malade, dès le 22 décembre, un pronostic fatal, si on ne l'entreprenait pas... Mes propositions furent rejetées... À partir de ce moment, on cessa de m'accorder autour de Gambetta la confiance dont j'avais joui jusqu'alors... » (1905 : 33)

En l'absence d'opération, Gambetta mourut le 31 décembre. Cette erreur de diagnostic et l'incapacité des médecins à prendre la bonne initiative a dans le jargon médical une appellation empreinte de délicatesse ; on appelle cela « abstention thérapeutique ».

L'immunologie – discipline à la base de la conception contemporaine de l'art médical – ne propose plus seulement des balises ponctuelles, « organicistes » et synchroniques, mais la trame complexe et diachronique de la causalité à l'œuvre dans la maladie. Le nouveau langage médical est celui du conte immunologique qui met en scène un scénario policier où l'intrigue est fournie par l'enchevêtrement logique et parfait d'une organisation de défense de l'organisme. C'est mon troisième paradoxe : le corps médical tient un discours connoté, affectif, dans la mesure où le rapport à la pathologie est humain. Dès lors, le discours médical dans son ensemble – car l'immunologie traverse horizontalement la structure physiologique – devient métaphorique et recèle, derrière un style en apparence impersonnel, un combat tout personnel contre la maladie. Pour le dire autrement, peu à peu le sujet pensant est réhabilité au sein de la communauté des patients.

Voici un exemple que l'on retrouve dans cet aphorisme classique de la chirurgie abdominale :

« Le ventre de la péritonite ne crie pas toujours au secours ».

Ou encore, toujours à propos de la péritonite, ce conseil pour la palpation :

« Continuez à parcourir tout le ventre. Il est à peu près uniformément douloureux, tendu, hyperesthésique. La douleur provoquée, son maximum ? Comme c'est difficile à préciser : partout, en effet, la défense musculaire instantanée, rigide, indépressible : c'est le *ventre de bois*. À force de patience, vous vous êtes cependant convaincu que c'est bien à droite et en bas qu'est la douleur la plus vive et la contracture la plus durcie ». (Mondor 1949 : 72)

Ainsi, je dirais que la quête des enjeux enfouis sous la surface linguistique déplace l'attention du traducteur vers l'auteur du texte et vers son destinataire.

Le lecteur non averti fera l'impasse sur ces enjeux latents ; mais le traducteur médical a pour tâche de faire émerger le propos caché et d'éventer la ruse implicite. En présence d'un texte médical, le traducteur se doit d'être un médecin des âmes. Comme je l'ai écrit ailleurs, un texte bien documenté n'est pas forcément un texte bien traduit (2001 : 92-102).

Ordonner ces notions secourables, non pas en une longue suite, mais dans leurs liaisons ordinaires et leurs écarts éventuels, m'a paru justifier l'effort de cet exposé.

BIBLIOGRAPHIE

- Balliu (C.), 2001 : « Les traducteurs : ces médecins légistes du texte » dans *Traduction médicale et documentation, Meta*, 46-1, pp. 92-102.
- Balliu (C.), 2002 : « Ça vous chatouille ou ça vous grattouille ? » dans *Les Maux en mots, Traduire*, revue de la SFT (Société française des Traducteurs), n° 194, pp. 77-101
- Balliu (C.), 2005 : « Traduire les maux » dans *Forum*, Presses de la Sorbonne Nouvelle et KSCI, vol. III, n° 1, pp. 1-19
- Charcot (J. M.), 1889 : *Leçons du Mardi*, Delahaye et Lecrosnier édit., Paris.
- Chibout (K.), Kotowicz (J.-P.), Briffault (X.), 2003 : « Systèmes collecticiels et communication naturelle : apports conjoints de l'informatique linguistique et de l'ergonomie » dans *Revue d'interaction homme-machine Journal of Human-Machine Interaction*, Europia, Paris, pp. 103.
- Jeay (A. M.), 1991 : *Les Messageries télématiques*, Eyrolles, Paris.
- Lannelongue (O.-M.), 1905 : *Cliniques chirurgicales*, Paris. On consultera également : *Blessure et maladie de Gambetta*, 1883.
- Mondor (H.), 1949 : *Diagnostics urgents-abdomen*, Masson, Paris, pp. 225.
- Navarro (F. A.), 1997 : Traducción y lenguaje en medicina, *Fundación Dr Antonio Esteve, Barcelona*, pp. 10.
- Ochanine (D.), 1978 : « Le rôle des images opératives dans la régulation des activités de travail » dans *Psychologie et éducation*, pp. 63-79.

L'INCIDENCE DU CONTEXTE SUR LA TRADUCTION MÉDICALE

Renate Trurnit Verbic

LILLA/CRDL

Université de Nice Sophia-Antipolis

jr.verbic@wanadoo.fr

Résumé : *L'incidence du contexte dans la compréhension et la traduction devient apparente en alignant des unités phraséologiques provenant de textes médicaux français et allemands. Cotexte et contexte restreignent successivement le champ d'application des unités et spécialisent leur sens. Lors de leur traduction, le choix lexical dans la langue cible en dépend. Il est donc important dans une base de donnée bilingue que des étiquettes contextuelles permettent de vérifier l'équivalence des contextes en langue source et en langue cible.*

Mots-clés : *traduction médicale, unités phraséologiques, français, allemand, contexte*

1. INTRODUCTION

La thématique du colloque *Mots, termes et contexte* semble offrir un prolongement idéal à notre thèse sur l'analyse de textes médicaux en français et en allemand dans le domaine de l'asthme : incidences sur la traduction¹. Nous y avons en effet étudié le fonctionnement de la langue médicale en vue de sa modélisation dans le cadre d'applications telles que la traduction assistée par ordinateur et dans une perspective bilingue, français <> allemand. Notre choix d'étudier des textes médicaux en français et en allemand à partir de la phraséologie propre à chaque langue avait été motivé par l'absence d'outils de traduction performants dans le domaine médical entre ces deux langues, alors que la demande de traduction dans le domaine médical ne cesse d'augmenter, comme l'a aussi souligné Hannelore Lee-Jahnke (2001 : 145) dans son article sur *L'enseignement de la traduction médicale : un double défi ?*

Les réflexions que nous vous proposons ici sur le contexte trouvent leur origine dans l'analyse du *corpus*² qui a servi de base de travail pour notre thèse. Cette analyse s'est déroulée par étapes. Ainsi l'analyse thématique des textes a-t-elle conduit à une description notionnelle du domaine, l'analyse du lexique dans les deux langues a abouti à la description notionnelle des lexiques français et allemand. L'analyse des paradigmes d'emploi des termes les plus fréquents a permis de trouver des unités phraséologiques (UP) françaises et allemandes du domaine ; le classement de ces paradigmes en classes d'emploi nous a per-

¹ Thèse soutenue en décembre 2004 (Université de Nice) et préparée sous la direction du professeur Henri Zinglé (LILLA-EA 2140, devenu depuis peu le CRDL : Centre de recherche en didactique des langues).

² *Corpus* d'UP : 5 673 enregistrements bi- ou trilingues, en anglais, en français et en allemand, dont 2 085 sont alignés entre le français et l'allemand, 672 sont alignés en français et anglais et 400 en anglais et en allemand, environ 400 unités sont alignées dans les trois langues. D'autres langues peuvent s'y ajouter.

mis d'aligner des unités de traduction dans un même contexte d'utilisation. Chaque étape a fourni des cadres successifs à l'intérieur desquels le sens d'une UP se trouvait précisé et pouvait être aligné avec une UP de la langue cible apparaissant dans un même contexte.

Le schéma ci-après visualise les étapes de l'analyse et son aboutissement :

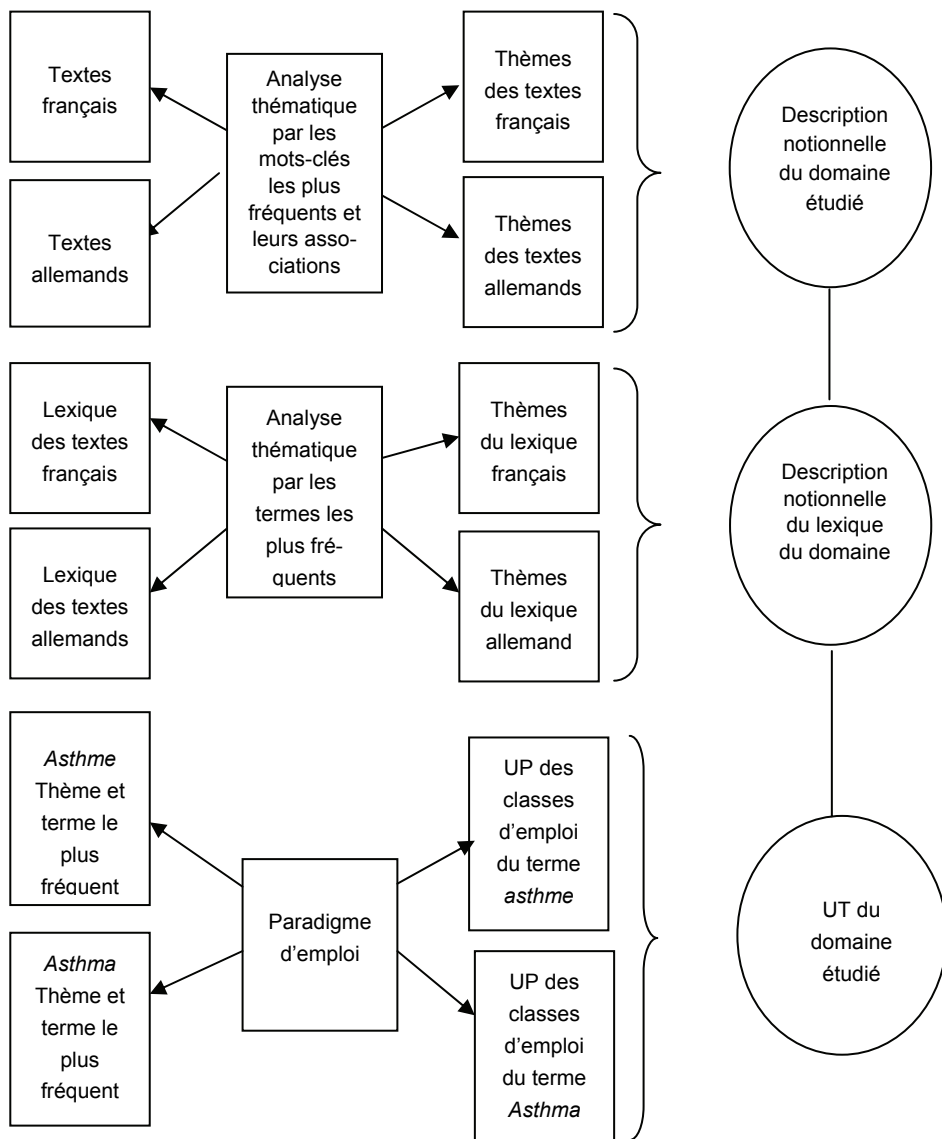


Tableau 1. : Schéma d'analyse du corpus d'étude

Lors de l'analyse du *corpus* et pendant la rédaction de notre thèse, le terme *contexte* nous a servi de nombreuses fois pour expliquer le choix d'une lexie dans la langue cible. Pour nous, il renvoyait alors soit au domaine spécialisé, soit à la nature et à l'intention du texte

médical, soit à une situation d'emploi ou à un usage à l'intérieur du domaine spécialisé. Il s'agissait toujours pour nous d'un paramètre sélecteur de sens, déterminant pour justifier la traduction. Mais qu'entend-on en général par le terme *contexte* ?

2. CONTEXTE : DE QUOI PARLE-T-ON ?

Le rôle du contexte dans la compréhension et l'interprétation d'une énonciation est reconnu depuis longtemps. De nombreux linguistes en ont parlé, parmi lesquels il convient de citer Chomsky (1965 : 97, 127 N, 129, 157, etc.) qui reconnaît en 1965, dans les *Aspects de la théorie syntaxique*, l'importance du contexte pour la génération automatique du langage, en introduisant une validation par le contexte pour ses règles de réécriture : *context sensitive* (règles dépendantes du contexte) ou *context free* (règles indépendantes du contexte). Searl (1979 : 171) souligne dans *Sens et Expression, études et théorie des actes de langage*, que « le sens littéral de la phrase est toujours relatif à des assomptions contextuelles » ; Minsky (1981 : 95-128), dans *A Framework for Representing Knowledge*, présente un système de compréhension automatique du langage, fondé sur la séparation des connaissances, dans des « micro-worlds » ou cadres cognitifs reliés entre eux ; il insiste sur la nécessité de prendre en compte le contexte pour comprendre le vrai sens d'un terme. Mel'čuk, Clas, Polguère (1995 : 9), dans *l'Introduction à la lexicologie explicative et combinatoire* (l'ILEC), reconnaissent qu'il est habituel en lexicologie de « laisser, dans l'interprétation du sens, un rôle vital au contexte et à l'intuition des locuteurs », même si, dans leur approche de formalisation du lexique, ils ont dû « opter pour une prise de décisions discrètes (dans le sens mathématique) et absolues » du sens lexical. François Rastier (1991 : 154), quant à lui, considère « le contexte comme l'ensemble des instructions contenues dans le texte qui permettent d'identifier un sémème et les traits qui le composent ». Le contexte est pour lui l'élément déterminant du sens. Pour la prise en compte du contexte en terminologie, Jacques Boissy (1995 : 43-47) et Henri Zinglé (1997 : 22) considèrent l'ensemble de la phrase comme contexte d'un mot. Enilde Faulstich (1997 : 26), enfin, différencie le co-texte du contexte, le co-texte se composant des éléments linguistiques pré- et postposés ou sous-entendus d'un mot, le contexte se composant des éléments extralinguistiques qui ont une incidence sur le sens d'un énoncé.

Le contexte est donc reconnu comme une exigence de l'environnement, qui influe sur l'organisation interne et externe de systèmes d'interaction de la langue et intervient dans toutes les disciplines linguistiques : grammaire, syntaxe, sémantique, pragmatique, intelligence artificielle et traduction. Nous allons essayer, à l'aide d'exemples, de montrer que le co-texte à l'intérieur d'un contexte a aussi une incidence sur la compréhension et la traduction dans le domaine de la médecine.

3. INCIDENCE DU CONTEXTE SUR LA COMPRÉHENSION DES TEXTES MÉDICAUX

Le terme médical étant par vocation considéré comme univoque, ce sont surtout les lexies de la langue générale qui prennent leur sens à l'intérieur de leur co-texte dans un certain contexte qui est, ici, médical.

3.1 Le mot dans son co-texte

Un verbe donné prend son sens en fonction des lexies grammaticalement associées. Ainsi le verbe *traiter* prendra-t-il un sens différent selon son domaine d'application et selon son argument. En médecine, le verbe *traiter* aura le sens de « soigner (avec des médicaments) » en association avec des termes appartenant au concept « malade, maladie » :

traiter Med. + [malade, maladie] → < soigner (avec des médicaments) > :

[1] traiter un malade [malade] ;

[2] traiter la BPCO [maladie].

Toujours dans le domaine médical, le verbe *traiter qqch.* associé au concept « exemplaire » impliquera le sens d'« examiner, analyser » :

traiter Med. + [exemplaire] → < examiner, analyser > :

[3] les échantillons [*exemplaire*] doivent être traités (dans les deux heures qui suivent l'expectoration).

Dans d'autres domaines, comme celui du bâtiment, des affaires ou de la politique, le verbe *traiter qqch.* pourra signifier « appliquer un produit de protection », « discuter en négociant qqch. » ou « soumettre qqch. à une discussion » et ce selon l'objet associé. Le co-texte immédiat du terme oriente l'interprétation du sens du verbe (*cf.* Verbic 02). Dans un contexte d'énonciation, le lecteur ou l'interlocuteur déduit le sens de son domaine d'application, ce qu'un système automatique ne peut faire, si aucun indice ne lui est donné. Pour permettre une reconnaissance automatique du sens, il faut introduire une étiquette liant le sens du verbe à une classe d'objet : « malade », « maladie » ou « exemplaire » par exemple, à l'intérieur du domaine de spécialisation : « médical ». Le verbe est ainsi lié à son contexte d'énonciation. Ces premières contraintes sémantiques sont les premiers indices dans « le parcours interprétatif » (Rastier 91 : 154) d'une unité phraséologique d'un domaine de spécialité.

3.2. Du co-texte au texte

Le domaine d'application et le co-texte immédiat ne suffisent pas toujours à éclairer le sens d'un énoncé. Ainsi, pour saisir le sens de la phrase « (il semble que) la présence d'éosinophiles activés est la principale caractéristique », faut-il élargir le co-texte au paragraphe entier pour comprendre que le mot *caractéristique* renvoie au terme *asthme* et qu'il s'agit d'une caractéristique de l'asthme³.

Ensuite, il faut encore situer le paragraphe dans le texte, un cours sur l'asthme, *Cours DCEMI* : « Asthme : Objectifs, texte et diapositives » et, à l'intérieur du texte, dans la thématique du chapitre qui traite de la *définition de l'asthme* et des Signes anatomopatholo-

³ P. ex. « Certains auteurs ont défini l'asthme comme une bronchite chronique desquamative à éosinophiles ; en fait il faut retenir qu'il existe une inflammation des voies aériennes, et que cette inflammation est polymorphe, diffuse et réversible ; mais il semble que la présence d'éosinophiles activés est la principale caractéristique ».

gies. C'est seulement alors que nous comprenons l'arrière-fond intellectuel sur lequel se construit le sens (Searl 79, 35 et ch. 5 : 178, 179, 184, 186), c'est-à-dire que « la présence d'éosinophiles activés est la principale caractéristique » de l'asthme et que c'est cette caractéristique-là qui permet de différencier l'asthme d'une autre maladie respiratoire, la BPCO (bronchite pulmonaire chronique obstructive), qui présente des symptômes similaires.

Notre « parcours interprétatif » s'est enrichi d'informations concernant la thématique des textes et des chapitres. Ces thèmes ont servi sous forme de mots-clés pour relier chaque enregistrement d'une unité phraséologique à son contexte d'utilisation. Ainsi l'UP :

[4] être la principale caractéristique de qqch.

est-elle reliée à une étiquette de trois mots-clés appartenant au domaine médical : [Med. : asthme, définition, anatomopathologie]. Ces mots-clés correspondent à la thématique principale du texte, à la thématique du chapitre et du sous-chapitre du texte source de l'UP. Elles permettent un premier classement des UP utilisées dans le domaine de l'asthme.

3.3. Texte et entour

Pour aller plus loin dans la mise en contexte, nous devons inclure des informations sur les auteurs et la publication de leurs articles. Prenons par exemple une phrase provenant d'un article d'Étienne Rica : « L'asthme et les allergies respiratoires », publié sur le site <http://www.frm.org/Scientifique/Sujetsfond/asthme/cadasthm.htm> en octobre 1997 et actualisé en octobre 2000. L'article avait été réalisé en collaboration avec Alain Grimfield, de l'hôpital Trousseau à Paris, et Francisque Leynadier, de l'hôpital Tenon à Paris.

La phrase extraite de cet article ne prend tout son sens qu'en analysant toutes les informations linguistiques et non linguistiques qui s'y rapportent :

[5] Bien que compliquée, cette stratégie semble efficace, au moins expérimentalement chez le lapin.

L'utilisation du verbe *sembler* permet de comprendre qu'il ne s'agit pas encore d'une certitude scientifique, mais d'une possibilité de recherche. La thématique du texte place cet énoncé dans le domaine de l'asthme et, à l'intérieur du texte, dans le chapitre « Les voies de recherche sur les nouveaux traitements ». Les noms des auteurs et la référence à leur lieu de travail permettent finalement de le situer dans un contexte de recherche hospitalière à Paris. La publication sur un site scientifique Internet indique un souci de partage des connaissances.

Pour vraiment comprendre un texte médical, il faut donc tenir compte de toutes les informations, linguistiques ou extralinguistiques, fournies par le lexique, par l'utilisation du lexique à l'intérieur d'un texte, par la fonction du texte à l'intérieur du domaine. La construction du sens se fait ainsi par étapes.

4. INCIDENCE DU CONTEXTE SUR LA TRADUCTION MÉDICALE

La compréhension du sens d'un énoncé permet sa traduction. Car, comme le disait déjà Hieronymus ou saint Jérôme, le saint patron des traducteurs, on ne traduit pas le mot mais

le sens du mot. Dans une lettre à son ami Pammacius, il explique sa technique de traduction des textes non bibliques : « *Libera voce profiteor, me in interpretatione Graecorum... non verbum e verbo, sed sensum exprimere de sensu*⁴ ». Nous avons vu que le sens se construisait par des informations successives apportées par le domaine d'application, par l'utilisation des termes ou mots à l'intérieur d'un co-texte ou d'un texte permettant sa mise en contexte. Afin de montrer l'incidence du contexte dans la traduction médicale, nous donnons pour chaque étape interprétative des exemples qui justifient la prise en compte du contexte dans la traduction médicale.

4.1. Domaine et traduction

Le domaine d'application est important pour la traduction. Il faut connaître l'équivalent dans la langue cible du terme utilisé dans les mêmes conditions d'application. Par exemple, l'équivalent allemand du terme *attaque* sera différent selon le contexte d'utilisation : dans le domaine militaire, le terme *attaque* sera traduit par *Angriff* ; dans le domaine policier, il sera traduit par *Überfall* et dans le domaine médical, par *Anfall*. Le terme *attaque d'asthme* devra donc être rendu par *Asthmaanfall* et non par *Asthmaangriff* ou *Asthmaüberfall*. Il faut donc situer un terme dans son domaine d'application pour trouver la traduction correcte dans la langue cible.

4.2. Co-texte et traduction

Dans le domaine médical, le terme allemand *Anfall* ne sera pas l'équivalent exclusif du terme français *attaque*. Sa traduction variera selon la combinatoire terminologique utilisée dans le domaine. *Anfall* sera traduit en français par : *attaque*, *quinte*, *crise* ou *accès*, selon la combinatoire terminologique propre au domaine ou au sous-domaine de la médecine :

- attaque d'asthme <> *Asthmaanfall* (domaine : asthme) ;
- quinte de toux <> *Hustenanfall* (domaine : maladies respiratoires, asthme) ;
- crise cardiaque <> *Herzanfall* (domaine : cardiologie et maladies respiratoires) ;
- accès de vertige <> *Schwindelanfall* (domaine : neurologie, etc.).

La bonne connaissance de la terminologie de la langue cible ne suffit pas toujours. Le traducteur doit aussi connaître l'emploi de ces termes à l'intérieur du domaine spécialisé, car « les langues de spécialité utilisent de façon particulière la langue générale comme support nécessaire à la communication » (cf. Boissy 95, p. 47). Selon le contexte d'usage dans une langue, un sème est exprimé ou sous-entendu. Ainsi, la relation entre le malade et sa maladie, qui sera exprimée, en français, de préférence par le verbe *souffrir*, le sera-t-elle, en allemand, par la simple préposition *mit* « avec », qui associe le malade à sa maladie. En allemand, le thème de la souffrance est souvent ignoré ou seulement sous-entendu dans la relation entre le malade et sa maladie :

[6] patients souffrant de rhinite/*Patienten mit Rhinitis* (litt. patients avec rhinite).

⁴ Saint Jérôme (né à Stridon entre 340 et 350 — 420) : « Je reconnais utiliser une parole libre pour la traduction du grec... non pas tirer un mot d'un mot, mais un sens d'un sens. »

Le bon choix des termes du domaine dépendra donc aussi des connaissances que le traducteur a de leur usage d'utilisation dans la langue cible.

4.3. Texte et traduction

Domaine et co-texte influent sur la traduction, tout comme le genre des textes à traduire. Un texte publié dans une revue scientifique n'utilisera pas le même vocabulaire qu'un texte de vulgarisation. Ainsi, dans un contexte de publication scientifique, le traducteur choisira, pour traduire *asthme cardiaque*, le terme *Asthma cardiale* et non *Herzasthma* (lit. asthme du cœur), qui n'appartient pas à la langue spécialisée et est plutôt utilisé dans des articles de vulgarisation. Cette référence au niveau de spécialisation est importante pour l'allemand médical, car celui-ci peut utiliser, selon le public visé, trois niveaux de langue. Le degré le plus spécialisé est une terminologie savante gréco-latine, dont les termes sont appelés *termini technici* (abréviation usuelle : *Termini*). Leur déclinaison se fait selon la grammaire latine. Ces termes sont les plus aptes, selon le dictionnaire spécialisé Duden (1998 : 26), à la communication écrite entre spécialistes, même si, dans les publications, on trouve surtout les *Trivialbezeichnungen* (lit. désignations ordinaires). Il s'agit d'une terminologie, toujours d'origine gréco-latine, mais dont l'orthographe et les déclinaisons ont été adaptées à l'allemand. En plus, il existe une terminologie d'origine allemande, destinée à la communication avec des non-spécialistes, les *volkstümliche Bezeichnungen* (litt. désignations populaires). Ces dernières, moins nombreuses, sont considérées comme l'échelon le plus bas des termes spécialisés :

Terminus	Trivialbezeichnung	Volkstümliche Bezeichnung
(noxa)	Noxe	<i>Schadstoff</i> (lit. substance nocive)
Cella pl.	<i>Zelle</i> (cellule)	---
Pneumonia	Pneumonie	<i>Lungenentzündung</i> (lit. inflammation des poumons)
Trachea	Trachee---	<i>Luftröhre</i> (lit. voie respiratoire)

Tableau 2. : Différents niveaux de spécialisation des termes allemands

La référence à la publication est donc importante pour situer le niveau de spécialisation pour la traduction.

Ces différents éléments, essentiels pour l'interprétation du sens et la traduction d'un mot, d'un terme ou d'une UP, doivent être pris en compte lors de la constitution d'une base de données. De ce fait, lors de la création de la nôtre, nous avons associé à chaque enregistrement d'une UP des informations d'ordre syntactico-sémantique et contextuel.

5. PRISE EN COMPTE DU CONTEXTE DANS UNE BASE DE DONNÉES D'UP

Pour notre *corpus* d'étude, nous avons constitué une base de données dans un formalisme compatible avec une exploitation des données par la *ZStation* : « un environnement de génie linguistique fondé sur les méthodes de l'intelligence artificielle distribuée et orientée vers la modélisation de connaissances linguistiques en vue d'applications liées au traitement automatique des langues » (Zinglé 1997 : 79). Un certain nombre d'applications, dérivées de cette approche, nous a servi pour analyser le *corpus* ; avec *ZLoc* nous avons trié, lemmatisé et compilé le *corpus* en base de données phraséologiques. Des fonctions permettent de créer des fichiers soit pour l'impression (format dictionnaire) soit en liaison avec un traitement de texte (*cf.* Zinglé 1999 : 282).

Pour le traitement automatique des données, chaque UP se présente accompagnée d'un bloc d'informations, d'ordre fonctionnel, sémantique et contextuel.

5.1. Étiquettes ou indices de sens

Les informations sont données sous des étiquettes identiques pour chaque enregistrement. Dans notre base de données monolingues, l'UP *souffrir de qqch.* se présentera comme suit :

[7] « souffrir de qqch. < asthme bronchique, phr. [Asthme, Traitement] ; déf. Être atteint douloureusement d'une forme de dyspnée caractérisée par une difficulté de l'expiration accompagnée d'un bruit sifflant. En langage clinique, asthme bronchique (asthme vrai ou essentiel). [Mas.] ; Ex. En Suisse, environ un enfant sur dix souffre d'asthme bronchique ; [réf. : *mps* (ath) : 16.11.00] ».

Ce bloc d'information contient les éléments suivants :

- unité phraséologique (UP) : souffrir de qqch < asthme bronchique > ;
- marque fonctionnelle : phr. (phrasème). L'UP est située dans sa fonction grammaticale et relationnelle ;
- mots-clés : [Asthme, Traitement]. L'article concerne l'asthme, l'unité provient d'un chapitre sur le traitement de l'asthme ;
- définition (déf.) : *cf.* texte plus haut. Aide à la compréhension de l'information médicale de l'UP. La référence [Mas.] indique que cette définition a été faite avec l'aide du dictionnaire médical édité par Masson (1992) ;
- exemple (ex.) : « En Suisse, environ un enfant sur dix souffre d'asthme bronchique. » L'unité phraséologique est montrée dans son environnement phrastique, qui montre son co-texte : le sujet associé est représenté par le mot *enfant*, et montre le contexte d'utilisation : l'UP renvoie à une situation en Suisse et à un % d'enfants ;
- référence du texte (réf.) : [*mps* (ath) : 16.11.00]. Cette référence renvoie au texte source de l'unité phraséologique : *Medpress* [*mps*] publié le 16.11.2000 à Bâle avec le titre : « Asthme bronchique : les filles sont moins souvent traitées que les garçons ».

Ces informations linguistiques et contextuelles permettent de situer une UP monolingue dans son contexte. Les étiquettes fonctionnelles (il en existe quatre : *phr.*, *sn.*, *adv.*, *qual.*) classent l'UP dans ses fonctions grammaticale et relationnelle, la définition ajoute la valeur sémantique et informationnelle, les exemples, les mots-clés et les références situent les unités dans leur co-texte et leur contexte de publication.

Ces étiquettes sont des indices de sens, qui ont également permis de relier chaque UP à trois concepts, dont le premier décrit la base ou l'« opérateur » (Boissy 95 : 43) de l'UP, le second la relation entre l'opérateur et son complément et le troisième le sens du complément. L'unité phraséologique : *souffrir de qqch.* < *asthme bronchique* > est ainsi représentée par deux concepts et une relation : [[subir] (association) [syndrome]]. Ce bloc conceptuel, qu'on pourrait comparer à l'« archi-concept » de Philippe Thoiron (1996 : 512-552), est une paraphrase simplifiée du sens et peut servir de pivot conceptuel pour la traduction. Le principe s'inspire pour nous des graphes conceptuels de Sowa (1984), sur lesquels s'appuie aussi l'analyse linguistique de la *ZStation* présentée plus haut.

Le niveau conceptuel étant celui de la traduction, le contexte intervient également sur le choix des concepts.

5.2. Contexte, un cadre dans lequel un concept prend une certaine valeur

À partir des concepts, nous avons pu trouver dans notre *corpus* allemand des UP équivalentes à ceux du *corpus* français. Ces UP allemandes avaient été collectées dans les mêmes conditions de départ que celles du *corpus* français, soit dans des textes médicaux du même domaine, d'un même niveau de langue, d'une même thématique et d'un même emploi.

Ainsi à partir des concepts liés à l'UP « souffrir de qqch. < asthme bronchique > », avons-nous pu regrouper les UP et les termes suivants :

[subir : fra = {être atteint de qqch. ; souffrir de qqch. ; ressentir les effets de qqch.}/deu = {*von etw betroffen sein, an etw leiden, (mit)*}];

[association : fra = {(de)}/deu : {mit}];

[syndrome : fra = {asthme, asthme d'effort ; asthme professionnel ; asthme bronchique ; asthme de faible gravité ; etc.}/deu = {*Asthma ; Belastungs-Asthma ; berufsbedingtes Asthma ; Asthma bronchiale ; leichtes Asthma ; etc.*}].

La valeur sémantique des concepts n'est garantie qu'à l'intérieur du contexte défini au départ, c'est-à-dire le domaine médical des maladies respiratoires, la thématique de l'asthme, la terminologie appartenant à cette thématique et qui est utilisée par des spécialistes du domaine, lors de leurs publications dans des revues scientifiques, à une époque définie. Jakobson disait déjà que la connaissance du contexte permet seule une traduction sans perte d'informations : « Plus le contexte d'un message est riche, et plus la perte d'information est limitée » (Jakobson 1963 : 84).

5.3. Contexte, le cadre pour valider une traduction

Si les concepts permettent la traduction des classes d'objets, la traduction ou l'alignement d'UP demande une mise en parallèle du contexte d'utilisation. Nous avons obtenu cette

mise en parallèle en analysant et en classant les paradigmes d'emploi des termes les plus fréquents dans les *corpus* français et allemand.

Ainsi les UP liées au terme asthme/*Asthma* ont-elles été groupées sous 20 thématiques d'emploi : *Types d'Asthme* (37)⁵, *Asthme et processus pathologique* (32), *Asthme et traitement* (32), *Asthme et son degré de gravité* (30), *Asthme et diagnostic* (27), *Asthme et études* (24), *Asthme et autres maladies* (24), *Définition de l'asthme* (18), *Asthme et sa caractérisation* (19), *Asthme et symptômes* (17), etc. À l'intérieur de ces classes, nous avons trouvé des expressions équivalentes ou unités de traduction (UT).

Nous donnons pour exemple quelques unités pour la classe d'emploi de la *définition* :

UT de la classe d'emploi de la définition		
	UP françaises / allemandes	Variable associée : concept et réalisation
UT	être qqch. \diamond	<[état_patho : (la) principale anomalie fonctionnelle de l'asthme]>
	<i>etw</i> <(N)> <i>sein</i>	<[état_patho : (die) wichtigste funktionale Abnormität bei Asthma]>
UT	il s'agit de qqch. \diamond	<[proc-pathol : (une) attaque d'asthme]>
	<i>es handelt sich um etw</i> <(ACC)>	<[maladie : (einen) Asthmaanfall]>
UT	définir qqch. comme étant qqch. \diamond	<[état_patho : (un) trouble inflammatoire chronique des voies respiratoires]>
	<i>etw wird definiert als etw</i> <(N)>	<[état_patho : (eine) chronische, entzündliche Störung der Atemwege]>
UT	avoir été défini comme qqch. \diamond	<[type de maladie : (un) asthme allergique]>
	<i>als etw</i> <(N)> <i>beschrieben worden sein</i>	<[maladie : allergisches A.b.]>
UT	affécter qqch. \diamond	<[loc. : (l') ensemble des voies respiratoires]>
	<i>etw</i> <(ACC)> <i>betreffen</i>	<[loc. : sämtliche Atemwege]>

Tableau 3. : UT de la définition

De même dans la classe du *processus pathologique*, de très nombreuses UT ont pu être trouvées, comme entre autres :

⁵ Entre parenthèses est donné le nombre d'UT dans une classe d'emploi du terme asthme/*Asthma*.

UT de la classe du processus pathologique		
	UP françaises / allemandes	Variable associée : concept et réalisation
UT	survenir sur <qqch. [maladie]>	<S. : ABPA> ; < [maladie] un asthme de type allergique >
	<i>bei <etw (D) [maladie]> auftreten</i>	<S. : ABPA> ; < [maladie] : <i>allergisches Asthma</i> >
UT	(S) [maladie] se constituer en quelques jours	<S.[maladie] : l'asthme aigu grave>
	(S) [maladie] <i>sich in einigen Tagen entwickeln</i>	<S.[maladie] : <i>akutes schweres Asthma</i> >
UT	(S) [maladie] évoluer typiquement de manière perannuelle	<S.[maladie] : L'asthme aux acariens>
	(S) [maladie] <i>sich typischerweise über das ganze Jahr (hin) entwickeln</i>	<S. [maladie] : <i>das durch Milben hervorgerufenes Asthma</i> >

Tableau 4. : UT de la classe d'emploi du processus pathologique

En dehors des conditions de départ pour la constitution du *corpus*, qui étaient d'utiliser des textes médicaux du même domaine, d'un même niveau de langue et d'une même thématique, nous avons introduit une information supplémentaire relative à la situation d'emploi d'un terme fréquent du domaine. Il me semble que les UT proposées dans une base de données qui donne, au-delà des informations fonctionnelles, sémantiques et conceptuelles, des informations sur le co-texte, le contexte de publication et surtout sur le contexte d'utilisation, peut garantir la validité d'une UT.

6. CONCLUSION

Nous espérons avoir pu démontrer que la sélection du vocabulaire dans une langue cible dépend d'informations multiples comme le domaine d'application d'un mot, d'un terme, son co-texte, mais aussi des informations sur le texte, sa thématique et son entour de publication : nom de la revue scientifique, nom de l'auteur, année de publication, public visé. Toutes ces informations permettent de tracer le cadre restrictif et indispensable qu'est le contexte. Il dirige le dernier choix du traducteur, il permet de choisir le mot juste, le terme juste, son juste emploi parmi d'autres. Il nous semble donc possible de prendre en compte le contexte dans la traduction médicale en indexant les entrées lexicales, non seulement par les informations fonctionnelles et sémantiques habituelles, mais en y ajoutant des informations contextuelles comme le domaine, la thématique du texte à l'intérieur du domaine, la référence au texte source, ainsi que la classe d'emploi d'un terme. Ces informations nous semblent suffisantes pour garantir un cadre de validité à une traduction. Dans un outil d'aide à la traduction, ces informations permettront au traducteur de trouver des termes ou les expressions correspondant à un même contexte d'application.

BIBLIOGRAPHIE

- Boissy (J.), 1995 : « Pour une analyse linguistique des termes dans la phrase » dans *Terminologies nouvelles*, Rint n° 14, pp. 43-47.
- Chomsky (N.), 1965, édition française 1971 : *Aspects de la théorie syntaxique*, Seuil, Paris.
- Coseriu (E.), 1994 : *Textlinguistik*, Coll. UTB für Wissenschaft, A. Francke Verlag, Tübingen und Basel.
- Delisle (J.), Woodworth (J.), 1995 : *Translators through History*, John Benjamins Publishing Company, Unesco Publishing, Amsterdam, Philadelphia.
- Dominicy (M.), 1984 : *La naissance de la grammaire moderne ; langage, logique et philosophie à Port-Royal*, Coll. « Philosophie et Langage », Mardaga, Bruxelles.
- Duden, 1971 : *Fremdwörterbuch*, Band 5, Bibliographisches Institut, Dudenverlag, Mannheim.
- Duden, 1998 : *Das Wörterbuch medizinischer Fachausdrücke*, Bibliographisches Institut et F.A. Brockhaus AG, Mannheim.
- Faulstich (E.), Rodrigues da Rocha (S. L.), 1997 : « A função pragmática do contexto lingüístico em obras lexicográficas e terminográficas » dans *Travaux du Lilla*, n° 2, Faculté des lettres arts et sciences humaines de l'Université de Nice Sophia Antipolis, Nice, pp. 23-32.
- Godard (P.) 1998-2000 : *Cours DCEM I* : « Asthme : Objectifs texte et diapositives » (date de création : 12 juillet 1998 – dernière mise à jour 09.09.00), <http://www.remcomp.com/asmanet/shanghai.html>
- Jakobson (R.), 1963 : *Essais de linguistique générale*, Les Éditions de Minuit, Paris.
- Jammal (A.), 1999 : « Une méthodologie de la traduction médicale », dans *Meta*, organe d'information et de recherche dans le domaine de la traduction, de la terminologie et de l'interprétation, Hommage à Jean-Paul Vinay (1910-1999), Les Presses de l'Université de Montréal, Canada.
- Lee-Jahnke (H.), 2001 : « L'enseignement de la traduction médicale : un double défi ? » dans *Meta*, organe d'information et de recherche dans les domaines de la traduction, de la terminologie et de l'interprétation, n° 46, 1, *Traduction médicale et documentation/Medical translation and documentation*, sous la direction du Dr Serge Quérin, Université de Montréal, Presses de l'Université de Montréal, Montréal, Canada, pp. 145-153.
- Meľčuk (I.), Clas (A.), Polguere (A.), 1995 : *Introduction à la lexicologie explicative et combinatoire*. coll. « Universités francophones », Duculot, Louvain.
- Minsky (M.), 1981 : « A Framework for Representing Knowledge » dans *Mind Design*, Ed. J. Haugeland, Cambridge, MA, The MIT Press, pp. 95-128.
- Rastier (F.), 1991 : *Sémantique et Recherches cognitives*, coll. « Formes sémiotiques », PUF, Paris.
- Rica (É.), 1997-2000 : *L'Asthme et les Allergies respiratoires*, réalisé en collaboration avec Grimfield (A.), (hôpital Trousseau, Paris), Leynadier (F.) (hôpital Tenon, Paris).
- Sabah (G.), 1988 : *L'Intelligence artificielle et le Langage*, t. I et II, Hermes, Paris.
- Searle (J. R.), 1979, Paris 1982 : *Sens et Expression : études de théorie des actes de langage*, coll. « Le sens commun », Les Éditions de Minuit, Paris.
- Sowa (J. F.), 1984 : *Conceptual Structures : Information processing in Mind and Machine*, IBM Systems Research Institute, Addison-Wesley publishing Company, Reading Massachusetts/Menlo Park, California/London/Amsterdam/Don Mills, Ontario/Sydney.
- Thoiron (Ph.), Arnaud (P.), Bejoint (H.) et Boisson (C. P.), 1996 : « Notion d'« archi-concept » et dénomination » dans *Meta*, Organe d'information et de recherche dans le domaine de la traduction, de la terminologie et de l'interprétation, vol. XLI, 4, Presses de l'Université de Montréal, Canada, pp. 512-524.

- Verbic (R.), 2002 : « Des paradigmes associés à l'unité phraséologique », dans *Travaux du Lilla*, n° 4, faculté des lettres, arts et sciences humaines de l'Université de Nice Sophia Antipolis, Nice, pp. 61-78.
- Zingle (H.), 1997 : « Outils et méthode d'extraction automatique de la terminologie » dans *Actes du Séminaire Realiter*, Nice, 1^{er} et 2 juillet 1996, Université de Nice, pp. 79-91.
- Zingle (H.), 1997 b : « Acquisition et traitement de données terminologiques avec *ZTermino* » dans *Travaux du Lilla*, n° 2, faculté des lettres, arts et sciences humaines de l'Université de Nice Sophia Antipolis, Nice, pp. 7-22.
- Zingle (H.), 1999 : « Terminologie et ingénierie linguistique » dans *Travaux du Lilla*, (juin 1999), *JILA '99*, Journées internationales de linguistiques appliquées, Nice, France 24 au 24 juin 1999, faculté des lettres, arts et sciences humaines de l'Université de Nice Sophia Antipolis, pp. 281-283.

L'EXPRESSION DE LA SUBJECTIVITÉ À TRAVERS LES VERBES DE PERCEPTION VISUELLE DANS LES TEXTES FRANÇAIS RELEVANT DES EXAMENS CLINIQUES ET PARACLINIQUES

Aneta Tosheva

section de langues étrangères
Université de médecine de Plovdiv
Plovdiv, Bulgarie
aneta_tosheva@yahoo.com

Résumé : *L'analyse de 15 textes français relevant d'examens médicaux montre que la subjectivité se glissant à travers les verbes de perception visuelle augmente dans les cas où l'interprétation cherche à éviter la conclusion définitive, ce qui impose une approche ethnolinguistique à la traduction vers le bulgare.*

Mots-clés : *traduction médicale, subjectivité*

1. INTRODUCTION. OBJECTIFS DE L'ÉTUDE

Les textes produits à partir des examens cliniques et paracliniques ou comptes-rendus d'examen (CRE) constituent une partie particulièrement importante du dossier médical et correspondent à un genre particulier au sein du discours médical.

Au cours de la traduction de ces textes du français vers le bulgare et inversement, nous avons constaté un taux élevé d'incidences d'asymétrie interlinguale. Cette asymétrie s'est fait sentir aux endroits où se laissait voir une fluctuation des limites entre l'objectif et le subjectif. Là où l'on s'attendait plutôt à des affirmations catégoriques provenant du caractère non dubitatif des examens, dans les textes français on a retrouvé une attitude verbale étayée par des moyens divers de modalisation subjective. Confrontées aux textes bulgares où la prise de position appréciative ou affective est bannie, ces productions textuelles viennent une fois de plus à l'encontre de la thèse préconisant que le discours spécialisé aurait une structure universelle reposant sur la standardisation du vocabulaire et les conventions des démarches descriptives interdisant toute appréciation ou prise de position affective.

Il nous semble donc légitime de procéder à l'éclairage des éléments créant des contextes différents au sein d'un même genre textuel et demandant des approches diversifiées de la part du traducteur médical.

2. MATÉRIEL ET DÉMARCHE À SUIVRE

Dans un *corpus* de 50 textes représentant des CRE, traduits du français en bulgare et commentés avec des spécialistes médicaux, nous avons repéré 15 textes représentatifs pour la manifestation de la subjectivité à travers l'expression de la modalité perceptive.

Considérant les textes comme des actes discursifs, nous allons définir leur contexte de production; ensuite, nous allons procéder à l'analyse du matériel lexical, verbes et constructions, traduisant la perception visuelle comme un moyen naturel de mettre en évidence le caractère objectif des faits observés. Nous allons y rechercher les traces de la subjectivité qui entraîne une déviation du mouvement argumentatif vers la non-précision, l'hypothèse ou la suggestion. Enfin, à la base d'une distinction typologique, nous allons identifier le caractère de la démarche à suivre lors de la traduction de ce type textuel.

3. CONTEXTE DE PRODUCTION DES TEXTES DE CR

Les textes étudiés sont des textes nécessairement écrits, rédigés par des médecins et adressés à un public relativement homogène. Formellement, ils sont représentatifs pour une situation communicative asymétrique, « non partagée » (Kerbrat-Orecchioni 1999 : 61), le destinataire du texte n'étant pas présent dans l'espace communicatif. Ils relèvent du discours formellement monologal mais à finalité dialogale et sollicitent une réponse différée en actes ou en paroles.

Le lieu de communication, qui est celui du discours officiel, affecte parfois l'aspect d'un courrier privé entre médecins; or, une convention institutionnalisée exige que tout échange épistolaire concernant le patient fasse partie intégrante du dossier médical de celui-ci. L'auteur du document est donc parfaitement conscient de ce que son message peut être lu par des personnes autres que le destinataire mentionné.

Les actants qu'il convient de prendre en considération dans cette analyse ont tous leur part à la construction de la subjectivité :

- *l'auteur-scripteur*, origine des relations interpersonnelles, qui est responsable des choix de sa manifestation explicite ou de son effacement de la trame discursive ;
- *le récepteur visé* qui intervient toujours indirectement, et la trace de sa présence virtuelle est, en dehors de l'indication nominative dans certains textes, « dans le choix de l'appareil stratégique, affectif ou argumentatif » [Kerbrat-Orecchioni 1999 : 160] du scripteur ;
- *le/le(s) récepteur(s) potentiel(s)* ayant accès au dossier médical ;
- *le regard médical expert*, l'œil clinique que nous mentionnons comme acteur autonome
- en dehors des sujets-actants, *le référent* du texte qui est une unité complexe.

D'un côté, ce sont les données réelles concrètement présentées dans l'espace communicatif et qui représentent le contenu du message. De l'autre côté, le dénoté effectif de ce message est le patient concerné dont le corps, ou l'organe, ou d'autres éléments matériels (prélèvements, cellules, coupes, etc.) font l'objet de l'examen réalisé en observation directe ou par l'intermédiaire d'une image.

Le but communicatif de ces textes, c'est l'exposition de faits observés et interprétés, ce qui définit une allure expositive et argumentative des séquences aboutissant à une conclusion engageant la personnalité et l'autorité professionnelle du sujet émetteur.

4. LA VISION ET L'IMAGE, LA PERCEPTION ET SON EXPRESSION VERBALE

Même si l'examen clinique fait appel au cinq sens, c'est *la vision* qui y occupe la place principale, fournissant l'information à l'observation experte aboutissant à la construction de *l'image*. Celle-ci relie les mondes du visible et de l'invisible par un jeu subtil d'analyse conjuguant sens, intelligence et savoir partagé. Cette image s'expose, s'impose à la vue et, à force d'exiger une interprétation immédiate, peut acquérir une autonomie qui se laisse facilement saisir par le langage. C'est la *perception*, à travers son expression linguistique, qui rend compte de cette primauté de l'image dans l'exposition objective des faits observés.

La perception correspond à une représentation intellectuelle où le sujet perceptif s'engage dans l'investigation et la connaissance d'un fait objectif, l'objet et/ou son image. Cette saisie de l'objectif implique un double parcours actif : une activité que l'objet exerce à l'égard du sujet en manifestant sa présence, et une activité du sujet perceptif intériorisant et analysant l'objet.

4.1. Un groupe important de verbes assurant la primauté de l'image perçue attribuent une activité accrue à l'objet d'observation [OO] en réduisant le sujet observant [SO] et producteur du texte à l'état d'observateur plus ou moins passif. C'est l'objet qui vient s'installer dans le champ visuel du sujet afin d'être intégré par l'activité cognitive de ce dernier avant de devenir l'objet de son interprétation.

Cette démarche conférant un caractère objectif à l'exposé est en fait le résultat d'un choix prémédité du sujet de minimiser son propre rôle d'observateur analysant et de laisser parler les faits eux-mêmes. Les verbes auxquels il a recours sont généralement pronominaux et assimilables aux tournures factitives *se faire voir, se laisser voir, s'offrir à la vue*, tels que *se présenter, se révéler comme, se montrer, se dessiner, se détacher, se découper, se déceler, se produire, se manifester, s'inscrire dans, se voir, s'apercevoir, se signaler, se constater*.

Dans les textes français étudiés, pas un seul emploi de ces verbes n'a été enregistré dans le contexte de la perception visuelle. On ne saurait interpréter ce fait autrement que par le sémantisme restreint de ces verbes réduits dans ce contexte à leur valeur strictement perceptive et par la tendance à la transitivity directe du français moderne.

Par contre, lorsqu'à l'idée de perception visuelle accordant une activité accrue de l'objet, se greffe l'idée d'intellection, le SO recourt à des verbes raffinant cette perception visuelle et précisant des détails spécifiques retenus lors de l'observation. Ainsi :

– la localisation et la disposition dans l'espace de l'observé est représentée par les verbes *se situer, se placer, se trouver, s'étendre ; constituer, contenir, abriter, comporter, comprendre, renfermer, s'intégrer à*, etc.

[1] Le chorion est abondant et abrite des glandes sans particularité.

– la mise en rapport entre l'objet et ses constituants ou d'autres objets est assurée par les verbes de correspondance – *intéresser, correspondre, répondre à, être lié à, se rattacher à* ; de coïncidence – *coïncider, couvrir, (se) recouper, recouvrir* ; de ressemblance – *ressembler à, se rapprocher de, différer de* ; de conformité – *convenir, contraster, etc.*

[2] Liquide brunâtre renfermant un agglutinat blanchâtre répondant à de la fibrine.

Dans l'idée de perception visuelle, l'OO et le SO sont complètement discriminés. « L'objet se laisse concevoir comme un projet objectif que le sujet, intervenant en deuxième lieu, intègre » (Manchev 1976 :19). Dans cette tendance à l'autonomie, l'objet peut se présenter comme une entité dynamique ou animée à travers les verbes et les expressions *émerger, surgir, former des îlots, rejoindre, coloniser, envahir, former saillie*.

[3] Succédant à la destruction épithéliale et provenant des zones périlésionnelles intactes, une migration cellulaire se développe après 3 à 4 jours d'évolution et vient coloniser les zones détruites.

Les constructions verbo-nominales qui abondent dans les textes étudiés mettent en valeur l'expression perceptive de l'idée de qualité qui est nominalisée. Ce sont *avoir, présenter, montrer, offrir, affecter... un aspect, une image, un air, une apparence, un caractère, une allure, une tendance ; (être) d'aspect, de forme... + hyperonyme ou nom présentatif + caractérisation ; les constructions participiales être entouré, cerné, enserré, centré, prolongé, tapissé, bordé, dépourvu, revêtu de, être représenté par, être visualisé, composé, constitué, mis en évidence, etc.* La ressemblance est évoquée toujours implicitement par rapport à la norme : l'OO *présentant la forme/l'image/l'aspect de...* est implicitement comparé à celui qui a *la forme/l'image/l'aspect* relevant de la norme. La raison de ces emplois devrait être recherchée dans la philosophie du regard expert du médecin : celui-ci ne voit pas les choses telles quelles, il les voit comparables à d'autres choses, c'est-à-dire douées de qualités que l'œil non expert ne saurait jamais identifier.

[4] Certaines zones montrent un tissu scléreux très collagène en voie d'ossification entouré par une réaction granulomateuse.

[5] Les corps vertébraux sus et sous-jacents du rachis thoracique et du sacrum présentent des remaniements d'allure cicatricielle.

4.2. À ce niveau de construction du sens où s'opère le choix lexical du SO, on peut déjà attester la présence d'une modalisation témoignant d'un degré élevé de subjectivation. Même si l'attribution d'une activité interne de l'observable confère au discours une allure expositive et objective, fondée sur les faits, le choix de s'effacer devant la réalité est un choix prémédité du sujet qui introduit des éléments de subjectivité pour faire état de ses incertitudes là où l'interprétation unique de l'image risque d'être compromise.

Quelques verbes n'appartenant pas au groupe 3.1. et qui accordent normalement une activité au sujet animé comme *voir, détecter, reconnaître* se sont convertis en adjectifs en *-able* afin d'imprimer un caractère actif à l'observé. Dans cet emploi, ces verbes sont assimilables à *être tel* tout en introduisant une dimension subjective. Il convient de mentionner ici l'emploi du verbe *apparaître* introduisant déjà une petite nuance de non-précision voire d'incertitude attachables à l'aspect observable de l'objet. Le cas limite entre l'objectif et le subjectif est signalé par le verbe *sembler* véhiculant la modalité d'attribution illusoire. Tout en laissant les qualités objectives de l'objet se manifester pleinement, le sujet prend ses distances quant à l'interprétation de cette objectivité en laissant percevoir un doute.

[6] Cette recoupe de 2,5 cm, examinée sur des plans de coupes apparaît dissociée.

[7] Catéthérisme difficile de l'orifice cervical externe qui semble sténose avec écoulement d'une hématométrie rétentionnelle.

4.3. Ayant été repéré et identifié après s'être offert à la vue, l'OO est soumis à l'interprétation du SO. L'interprétation de l'image est un parcours intellectuel actif dont les étapes sont complexes et demandent une formation experte.

– Dans les textes étudiés, le sujet observant et interprétant rapporte le résultat de son attitude perceptive active par l'intermédiaire des verbes *voir, percevoir, observer, retrouver, remarquer, noter*. L'insistance dans l'observation est marquée par *inspecter, étudier, explorer*. Une valeur intellectuelle y est infiltrée par les verbes *examiner, analyser*.

[8] À la dose de 2 MU d'interféron, nous avons observé au cours des derniers mois plusieurs épisodes de purpura des membres inférieurs et parallèlement une réascension des transaminases.

– La précision assurée à laquelle s'ajoute une activité intellectuelle d'interprétation et de commentaire implicite est représentée par les verbes *reconnaître, identifier, constater, trouver, retrouver, découvrir, déceler, dépister, détecter* ou substantifs déverbaux assimilables à ces verbes. On pourrait y classer le verbe *objectiver* et sa transcription analytique *objectiver la présence de*.

[9] Sur certains plans de coupes, on identifie une prolifération répondant à un carcinome épidermoïde invasif fait de massifs riches en atypies, certains centrés par des foyers de nécrose avec afflux de polynucléaires.

– Lorsque la perception visuelle résultant d'un travail de recherche exploratrice aboutit à une image apparue avec plus ou moins d'imprécision, ce sont les verbes *repérer, distinguer, discerner, visualiser, entrevoir* qui interviennent.

[10] L'origine du tronc coeliaque a été très difficile à visualiser en raison de la superposition artérielle et malgré la réalisation de deux séries de profil, l'une d'aortographie, la seconde après cathétérisme sélectif du tronc coellaque et reflux.

Le regard explorateur d'un SO actif est doué de pouvoir métaphorique. Il ne s'agit pas ici de cette manifestation banalisée de la métaphorisation qui caractérise le jargon médical de nos jours [Une angine qui vomit est une scarlatine] mais d'une transposition de traits perçus au cours de l'observation directe.

[11] Après cyto centrifugation, on identifie au sein de nappes denses d'hématies des lymphocytes et polynucléaires et des cellules mésothéliales parfois d'aspect irrité regroupées en petits îlots.

5. LE REGARD MÉDICAL ET LES LIEUX D'INSCRIPTION DE LA SUBJECTIVITÉ

Quels sont les rapports entre le processus de perception, sa représentation linguistique et l'image obtenue, dans l'optique de la subjectivité ? Référons-nous à un radiologue, Dr. Pierre Aimard. Dans son excellent exposé sur les enjeux de l'imagerie moderne, il affirme : « À l'encontre de ce que fit Narcisse, les médecins ne doivent pas se laisser séduire par ce

qui n'est qu'un reflet. Et pour paraphraser Magritte, on pourrait dire, en montrant une radiographie du thorax : Ceci n'est pas un poumon. Les images que nous fabriquons à partir du corps humain ne sont que des reflets de la réalité... » (Aimard, 1999).

En effet, l'OO interprétable comme un signe pathologique n'est qu'une image du réel, le signifiant dont la perception visuelle doit identifier le signifié. La relation SO/OO se trouve transposée dans le domaine de la relation entre le signe et le sens grâce à un opérateur, « le regard médical » (Foucault 1983). Celui-là, instrument et médiateur, renvoie chaque fois à un espace construit et modelé par un sujet actif et conscient. Les questions relevant des rapports entre signes naturels et signes langagiers, leur lecture et leur interprétation sont au fondement même de la communication instaurée à partir de ce sujet. Ses stratégies discursives apparaissent comme des stratégies d'investissement d'une forme d'un contenu, comme des démarches de partage.

Les verbes de perception que nous avons repérés dans le point 4 constituent l'appareil formel de cet instrument, de cet opérateur mariant le visuel à l'intuitif qu'est le regard médical. L'auteur-scripteur du texte y recourt pour privilégier soit le sujet perceptif, soit l'objet de perception. En dehors de ce choix qu'il entreprend dès le niveau lexical, il procède à d'autres choix lui permettant d'explicitier ou d'effacer sa présence dans le texte.

Ces choix définissent le taux de subjectivité dont les textes faisant l'objet de cette étude sont investis. Chercher ses lieux d'inscription, c'est rechercher toutes les traces linguistiques de la présence de l'émetteur liées à son activité perceptive.

– Ainsi, la première inscription personnelle de l'auteur dans son discours c'est le déictique *je*, beaucoup plus subjectif que le pluriel *nous* ou l'indéfini *on*. Ces derniers marquent le recul du sujet scripteur derrière une personne collective ou indéfinie. Or, le pronom *je* ne s'associe jamais dans l'ensemble des textes à un verbe perceptif, à l'exception du verbe *voir* à la valeur de *rencontrer* : *Je viens de voir à ma consultation M^{me} X*. Par contre, les formes verbales sont régulièrement précédées de *on*, et plus rarement de *nous*, cf. :

[12] On y trouve également des foyers macrophagiques et giganto-cellulaires.

[13] À la dose de 2 MU d'interféron, nous avons observé au cours des derniers mois plusieurs épisodes de purpura des membres inférieurs et parallèlement une réascension des transaminases.

Divers éléments, expressions, tournures ou structures grammaticales, prêtent au message une orientation argumentative. Il en découle une imbrication des dimensions descriptive et argumentative dépendant exclusivement des choix opérés par l'auteur. Ceux-ci sont conditionnés et limités par les contraintes relevant des données situationnelles et les conventions du genre.

– La subjectivité est saisie à travers des lexèmes qui énoncent *a priori* un jugement de valeur et un engagement émotionnel du locuteur à l'égard du contenu de son message ou qui, appartenant aux autres registres du langage, tranchent sur l'environnement linguistique et sont facilement perçus par l'intuition du récepteur. Ce sont des adjectifs, adverbes et substantifs à charge stylistique appréciative, affective, dubitative ou concessive :

[14] Mi-mars, elle part en classe de neige sous ttt par josacine à cause d'une bonne trachéite.

– Divers procédés d'expression d'hypothèses, ou éventualités qui restent à être approuvées, sont mis en jeu par l'auteur :

[15] L'éventualité d'une fistule n'est pas éliminée mais l'orifice sigmoïdien n'a pas été mis en évidence.

[16] Il est possible que ce syndrome dépressif soit lié en partie à la reprise du travail récent.

– L'explicitation de l'approximation et de l'hésitation relèvent de la volonté du sujet observant de ne pas fournir une conclusion catégorique. On peut y mentionner les verbes relevant de la modalité illusoire, la construction verbale *pouvoir* + *verbe*, certains adverbes ou expressions vecteurs d'incertitude ou imprécision :

[17] Les différents segments médullaires semblent respectés.

[18] Indication : Coloscopie dans le cadre de la recherche d'une cause de polymyosite pouvant être intégrée dans le cadre d'un syndrome para-néoplasique.

[19] Cet aspect est assez superposable à celui de la biopsie précédente.

6. LES STRATÉGIES DU SUJET ÉMETTEUR

L'amalgame des éléments vecteurs de la subjectivité définit des stratégies fondées sur des formules modalisatrices complexes développant une argumentation nuancée. La tension argumentative de tous ces textes vise soit à conduire le récepteur à une conclusion quant au diagnostic ou à la démarche à suivre, soit à l'en détourner en lui offrant le choix d'en tirer ses propres conclusions ou d'entreprendre de nouveaux examens, ce qui nous amène à identifier trois stratégies :

6.1 Stratégies d'effacement

Si les pronoms *on* et *nous* sont généralement admis comme marquant l'impersonnalité due à la volonté de l'auteur de laisser parler les faits, leur combinaison dans une séquence avec un sujet inanimé s'appropriant la fonction d'agent, évoque déjà un choix prémédité de s'effacer de la trame discursive, *cf.* :

[20] Cliniquement on note effectivement une tumeur qui est polaire inférieure, qui se montre tout à fait lisse, qui n'a pas de caractère cliniquement suspect et qui est indolore.

Le remplacement du sujet émetteur en début de phrase par un sujet inanimé introduisant le cadre situationnel insiste sur le côté instrumental de l'examen. On y retrouve *examen, observation, inspection, revue, exploration, analyse*, mais c'est encore plus marqué comme stratégie d'effacement lorsque la fonction sujet est occupée par le nom *patient* :

[21] La biopsie du 2 mars 2000 avait conclu à une hépatite chronique C.

[22] Ce patient avait eu une endoscopie le 28 juin 1996 qui avait trouvé seulement une oesophagite érosive et une candidose.

Le choix d'un pronom pluriel ou impersonnel, d'un nom d'être inanimé pour sujet grammatical et de verbes du groupe 3.1. préfigure en quelque sorte le développement de ce qui suit : l'argumentation repose exclusivement sur les données obtenues des appareils et est présentée comme dépendante exclusivement des faits vus et interprétés tels quels. L'auteur-scripteur construit un cadre argumentatif qui semble définitif et qui est souvent soutenu par l'argumentatif *donc* :

[23] L'hypothèse d'une tumeur du côlon est donc presque complètement écartée grâce à cette coloscopie. Donc très faible probabilité d'un cancer du côlon.

La stratégie de l'effacement est souvent doublée d'une mise en valeur du savoir partagé avec le destinataire, ce qui conduit l'auteur à utiliser des procédés d'économie linguistique et relevant du présupposé. Ici, il s'agit non seulement d'une compétence linguistique, mais avant tout d'une sorte de métacompétence présupposant la capacité de décoder des faits déjà intériorisés par l'expérience, de connaître les règles et les algorithmes des démarches.

Parmi ces procédés, on pourrait mentionner la reconversion des verbes ou de constructions verbales en noms : *présence, diminution, augmentation*, et la négativation du discours *Absence de... Il n'y a pas de... Il n'existe pas de... On n'observe pas de...*

[24] Présence d'une formation à centre hypodense au niveau du Douglas.

[25] Aucune anomalie duodénale et aucune lésion érosive oeso-gastro-duodénale.

6.2. Stratégies de prise de distance par rapport à l'interprétation de l'image qui s'impose objectivement

Cette stratégie communicative se définit comme un ensemble de procédés associant les nuances dubitatives et hypothétiques de l'exposé à une prise de position personnelle à l'égard de l'interprétation

– Les nuances dubitatives et hypothétiques sont souvent accentuées ou doublées par des éléments modalisateurs mettant en valeur le recul de l'auteur face aux données à interpréter :

[26] Il semble bien exister une sténose relativement serrée à l'origine du tronc coeliaque avec une image d'empreinte supérieure pouvant correspondre à un ligament arqué.

– La posologie et les démarches à suivre sont présentées sous la forme de suggestions précautionneuses.

[27] Il serait intéressant de prendre l'avis d'un dentiste, voire d'un stomatologue pour la conduite à tenir. Le Prof. ... pourrait éventuellement voir cet enfant en consultation si vous le souhaitiez.

– L'association de plusieurs éléments véhiculant l'idée d'hésitation – concessifs, adversatifs, verbes modaux ou formules appropriées – détournent le destinataire d'une conclusion définitive :

[28] L'œsophage est normal mais il existe çà et là quelques petits éléments blanchâtres surélevés pouvant faire penser (en l'absence de prémédication par un comprimé blanc) à une mycose.

[29] Pour ma part, je pense que pour l'instant, la situation est à peu près stable.

– L'étoffement verbal ou nominal crée parfois l'impression que l'auteur aimerait interpréter les faits avec prudence, en contournant la formulation par trop directe du problème ou de la conclusion. Il en résulte parfois des tautologies :

[30] Sur cette coupe, on note la présence d'une formation tissulaire droite para-utérine, située en arrière des vaisseaux et en avant de l'uretère, évoquant la présence d'adénopathies.

La stratégie de prise de distance marque un effritement du parcours argumentatif. La volonté de l'émetteur de laisser un champ libre au récepteur pour une interprétation plurivalente le conduit à combiner plusieurs éléments et d'utiliser des formules complexes en faveur d'une conclusion qu'il juge encore discutable.

[31] Conclusion : Côlon assez mal préparé surtout à droite. Un polype peut avoir échappé en raison de la préparation insuffisante mais un gros cancer aurait été vu. Donc très faible probabilité d'un cancer du côlon.

6.3. Stratégies de mise en valeur de l'attitude affective du médecin vis-à-vis de l'état de son patient

Dans leurs écrits, les auteurs de ces CR ne manquent pas de faire sentir leur engagement humain et profondément personnel vis-à-vis du patient concerné et de l'allure de son traitement. Cette attitude affectionnée se montre en dehors et indépendamment de toute argumentation professionnelle.

[32] Je vois ce jour en consultation M. X qui présente malheureusement une évolution de sa maladie de Hodgkin avec surtout une évolution au niveau du rachis dorsal localisée au niveau de D7 à D10.

[33] Je vous confie en urgence M^{lle} X âgée de 10 ans [...] Son cas m'intrigue [...] Je me demande ce qui lui arrive.

On peut donc conclure que la subjectivité du texte médical français vu dans l'optique du visuel, recouvre des valeurs intellectuelles et émotionnelles desservant de manière raffinée l'acheminement argumentatif de la pensée. Elle débouche sur des prises de positions personnelles de la part du sujet-scripteur témoignant de son engagement humain et professionnel, notamment la distance à l'égard de ses propres conclusions et la prise en charge affective de l'état du patient.

Certains auteurs (Régent 1992 :73) considèrent que cette impression de manque de conviction est laissée par le souci de l'auteur « de se prémunir contre les contradictions et les réfutations qui ne manqueront pas de venir » et par leur crainte « d'outrepasser leurs droits ». Nous sommes plutôt tentés d'y voir un souci de préserver la face du destinataire dans le cadre des rituels de politesse entre collègues. Le profil du destinataire du texte influence lui

aussi le choix stratégique. Lorsque les spécialistes médicaux s'engagent dans un discours intradisciplinaire, ils semblent se garder bien de formuler leurs conclusions comme définitives et leurs interprétations sont avancées prudemment, sous forme d'hypothèses.

7. TYPOLOGIE DES TEXTES D'EXAMENS CLINIQUES DU POINT DE VUE DE LA SUBJECTIVITÉ QUI S'EN DÉGAGE

Les textes étudiés, tout en présentant la description d'une image qui se veut objective, ont nécessairement une visée argumentative associée à un mouvement conclusif. Ces trois visées, descriptive, argumentative et conclusive, se développent sur trois paliers toujours présents dans la structure macro-discursive des textes :

- a) la perception visuelle met en jeu le regard expert du médecin
- b) l'observation identifie une image dont les caractéristiques sont verbalisées
- c) l'interprétation crée des arguments qui amènent à une conclusion

Ce type textuel commence donc comme objectif, mais à c) il est censé se scinder en trois sous-types :

- A. textes objectifs, faisant état uniquement des données obtenues et présentant une conclusion ferme issue d'une argumentation solide
- B. textes arborant une subjectivité modérée traduite par des stratégies d'effacement et de modalisation dissolvant l'argumentation
- C. textes montrant un taux élevé de subjectivité associant des procédés de modalisation et des marques d'affectivité où le mouvement argumentatif est doublé d'émotivité

C'est la stratégie d'effacement et les textes de type A qui prédomine dans l'ensemble des CR bulgares. Le discours médical y apparaît très objectivé, distancié et aboutit naturellement à une conclusion fondée sur une argumentation solidement développée. Le traducteur se trouve confronté à une situation renvoyant à celle des formules de politesse – ainsi, la phrase «Veuillez agréer, Cher Monsieur, l'expression de ma considération distinguée» ne devra être transposée en bulgare que par *S uvajenie*, (=Avec respect.)

Les procédés de modalisation appellent donc des transformations et des reconversions au cours de la traduction en bulgare. Les marques émotives et affectives empruntées en apparence à d'autres registres langagiers, mais correspondant parfaitement en réalité à l'esprit d'analyse français, demandent des explicitations et des étoffement supplémentaires dans les textes d'arrivée bulgares.

Conscient de ce fait, le traducteur sensibilisé devrait être très prudent en traduisant du français les textes de type B et C. Il devrait bien se garder de présenter le sens naturel de l'auteur français de préserver la face du récepteur comme une hésitation sur le plan professionnel et devrait mettre en valeur les marques d'affectivité et d'engagement personnel voire émotionnel, généralement absentes dans les CR bulgares, comme des caractéristiques ethnolinguistiques.

BIBLIOGRAPHIE

- Aimard (P.) : « Imagerie médicale : les arcanes d'un succès », exposé lors de la conférence *L'Image*, le 16/9/1995, relu et révisé en février 1999 (web).
- Foucault (M.), 1983 : *Naissance de la clinique. Archéologie du regard médical*, Paris, PUF.
- Kaménova (R.), Mantchéva (A.) et Bechkova (R.), 1986 : *Cours systématique de langue française, partie constructive 2*, 1^{re} édition, Naouka i Izkoustvo, Sofia.
- Kerbrat-Orecchioni (C.), 1999: *L'Énonciation. De la subjectivité dans le langage*, 4^e édition, Paris, Armand Colin.
- Manchev (K.), 1976 : *Morphologie française*, 1^{re} édit., Sofia, Naouka i Izkoustvo.
- Régent (O.), 1992: « Pratiques de communication en médecine : contextes anglais et français » dans *Langages*, mars 1992, n° 105, pp. 66-75.

CONTEXTES ET NÉOLOGIE TERMINOLOGIQUE DANS LE DOMAINE MÉDICAL

Maria Teresa Rijo F. Lino

centro de linguística da Universidade Nova de Lisboa

U. I. Lexicologie, lexicographie et terminologie

Lisbonne, Portugal

unl.tlino@mail.telepac.pt

Résumé : *La linguistique de corpus, la terminologie textuelle, la socioterminologie ont donné une contribution très importante à la théorie de la terminologie. Ces perspectives ont eu des répercussions sur les nouvelles approches des corpus. L'évolution des concepts scientifiques se manifeste en discours dans les situations de communication spécialisées écrites et orales. Les néologismes sémantiques par polysémisation, extraits à l'aide d'outils informatiques, d'un corpus de spécialité thématiquement parallèle, dans le domaine de la médecine, rendent compte des nouvelles propriétés cognitives du concept scientifique.*

Mots-clés : *terminologie textuelle, corpus textuel informatisé de spécialité, néologie sémantique, polysémie*

1. La thématique choisie pour ces Journées scientifiques du réseau LTT, « Mots, termes et contextes », nous conduit à réfléchir sur la problématique des conceptions de *corpus*, de contexte et de néologie, qui, ces dernières années, constituent l'objet de notre recherche.

La linguistique de *corpus* ainsi que les perspectives récentes de la terminologie textuelle et de la socioterminologie ont donné une contribution très importante à la théorie de la terminologie. Cependant, aujourd'hui, la recherche en terminologie fait évoluer le concept de *corpus*, dont les principes de constitution doivent répondre à la rénovation des méthodologies en terminologie et en lexicographie de spécialité.

Nous avons créé en 1991, à l'Universidade Nova de Lisboa, un *corpus textuel informatisé*, dont la plus grande partie est constituée par des textes de langues de spécialité de plusieurs domaines ; c'est un *corpus informatisé de textes de spécialité*, un *corpus ouvert*, utile aux recherches en terminologie et en lexicographie de spécialité qui utilisent en parallèle les matériaux issus soit des bases de terminologie soit des bases de néologismes scientifiques.

Un *corpus* n'est ni un ensemble d'énoncés ni une archive de textes ; il se distingue d'une base textuelle, organisée sans objectifs précis. La construction d'un *corpus* « joue un rôle souvent crucial dans une analyse à partir de/en *corpus*. Il ne faudrait pour autant en déduire qu'une fois l'objectif de l'étude clarifié, il ne reste plus qu'à trouver les textes pertinents pour la mener à bien. [...] Cette notion de pertinence continue à évoluer tout au long de l'analyse » (Condamines 2005). Un *corpus* est un « regroupement structuré de textes inté-

graux, [...], éventuellement enrichis par des étiquetages, et rassemblés [...] de manière théorique réflexive » (Rastier 2005).

Un ensemble de critères sémantiques, lexicaux, pragmatiques et informatiques est sous-jacent à l'organisation du *corpus textuel informatisé de spécialité*, qui intègre des textes de plusieurs niveaux de spécialisation, de différents types de locuteurs et de situations de communication spécialisée écrite et orale. L'établissement de « critères de sélection de typologies de types, de genres de textes et/ou de discours [...] car, ce n'est qu'en fonction de l'établissement de ces critères qu'il est possible de garantir l'adéquation des textes aux objectifs préétablis » (Costa 2001) au travail de recherche en terminologie.

Par ailleurs, ce *corpus* présente les caractéristiques suivantes : il s'agit d'un *corpus monolingue, bilingue* ou *plurilingue, corpus de deux variantes* de la langue portugaise (portugais du Portugal/portugais du Brésil, du Mozambique, d'Angola, de la Guinée-Bissau, du Cap-Vert) ; un des objectifs de ce *corpus* de textes relatifs aux deux variantes ou normes est l'observation de la néologie scientifique et technique, en particulier dans le domaine des sciences de la santé.

Ce *corpus* textuel intègre un *corpus de spécialité de référence* constitué par des textes très récents de domaines scientifiques particuliers (la sénologie, l'imagerie médicale, entre autres). En parallèle, nous avons un *corpus lexicographique de spécialité* caractérisé par de nouvelles macro et micro-structures ; c'est le cas du *Dicionário Terminológico de Senologia* (Garcia 2000) dont la nomenclature ainsi que les définitions rédigées par des médecins sénologues et par des linguistes sont riches en néologismes scientifiques.

Comme les *corpus textuels de spécialité* vieillissent rapidement en raison de l'évolution du concept scientifique, nous utilisons des *corpus oraux de spécialité* recueillis auprès des spécialistes, en situation de communication orale spécialisée (dans le domaine de la sénologie, entre autres). Les *corpus* oraux sont, en général, plus difficiles à organiser et à collecter, en particulier dans le cas des langues africaines qui n'ont pas une graphie stabilisée.

Le concept de *corpus de spécialité* évolue de façon à permettre la description des particularités de l'unité lexicale spécialisée ; nous retenons les phénomènes suivants : a) la stabilisation du sens du terme ; b) les divers types de variation terminologique ; c) la néologie scientifique qui participe souvent aux phénomènes de variation.

Ces *corpus* sont gérés par des hypertextes et par d'autres logiciels : un concordancier¹, un logiciel d'alignement² pour les *corpus* parallèles bilingues et plurilingues et un extracteur de termes, l'*ExtracTerm*, élaboré par Rute Costa (2001) ; ce dernier est constitué par trois composantes : 1) l'*Etiqualex*, un dictionnaire d'étiquettes qui attribue ces étiquettes automatiquement sur les *corpus* ; 2) un logiciel d'analyse morphosyntaxique ; 3) l'*extracteur* de termes.

¹ Ce concordancier a été élaboré par Gabriel Lopes du Centre d'intelligence artificielle de l'universidade Nova de Lisboa. <http://di.fct.unl.pt>

² Ce logiciel a été conçu par Gabriel Lopes de façon à aider la réalisation des dictionnaires électroniques bilingues (Portugais/Français, Portugais/Anglais) pour le système de traduction automatique Systran, dans le cadre du projet européen Tradaut. <http://terra.di.fct.unl.pt>

2. Ces *corpus* remplissent plusieurs fonctions : l'analyse, l'attestation de données, et l'extraction, en particulier, de termes simples et complexes, de collocations et de phraséologies. Ils rendent possible la sélection de nouveaux termes et de néologismes scientifiques, l'observation de l'implantation des néonymes durant une certaine période, en particulier dans quelques spécialités médicales : la médecine interne, l'imagerie médicale, l'anatomie radiologique, la sénologie.

3. Plusieurs types de phénomènes de variation terminologique peuvent être analysés à partir des contextes extraits de ces *corpus* ; nous soulignons l'importance des variantes phonomorphosyntaxiques et des variantes relatives aux normes de la langue portugaise au Portugal, au Brésil et aux pays de langue officielle portugaise. La polysémie, la synonymie, la néologie sémantique, la néologie sémantique par polysémisation, en particulier dans les domaines scientifiques mentionnés ci-dessus.

Les nouveaux concepts se traduisent sur le plan linguistique par des néologismes scientifiques. Dans cette communication, nous avons retenu quelques particularités de la néologie sémantique. La métaphore et la métonymie sont des processus très importants, mais elles n'épuisent pas toutes les possibilités ; d'autres processus sont à l'origine de la néologie sémantique.

Les nouveaux aspects conceptuels, sur le plan linguistique, se manifestent à travers des néologismes sémantiques créés par des polysémisations de différents types. La polysémie nominale est le processus le plus fréquent.

La polysémie, comme soulignent I. Mel'Čuk, A. Clas et A. Polguère (1995) « est un fait linguistique de base – un trait fort répandu et important de toutes les langues naturelles ». Ce concept linguistique est au cœur même des recherches en sémantique d'un grand nombre d'auteurs³.

La polysémie est un phénomène régulier ; elle n'est pas un accident, mais une composante sémantique constitutive des langues naturelles.

Il y a un certain consensus sur la définition de polysémie : « 1) une pluralité de sens liée à une seule forme ; 2) une pluralité de sens non disjointe, mais unis par telle ou telle relation » (Kleiber 2005).

Toutefois, nous pouvons nous demander comment l'identifier et la délimiter dans les contextes de spécialité écrits ou oraux : 1) quand y a-t-il sens multiple d'un terme ? ; 2) qu'entendons-nous par sens qui ne soient pas disjoints relativement à une unité terminologique ?

Il y a polysémie lorsqu'il y a « existence de sens différents pour un même mot, mais des sens perçus comme reliés d'une manière ou d'une autre » (Cadiot et Habert 1997).

³ Cf. Cadiot (P.), Cruse (A.), Fuchs (C.), Gross (G.), Kleiber (G.), Martin (R.), Mejri (S.), Mel'čuk (I.), Clas (A.), Polguère (A.), Rastier (F.), Victorri (B.), entre autres.

Selon C. Fuchs (1996), « la polysémie comme beaucoup d'autres termes linguistiques, est une notion qui se laisse très facilement appréhender de manière intuitive, mais qui se révèle beaucoup trop rétive à une définition rigoureuse ».

Plusieurs problématiques se soulèvent à propos de ce phénomène : d'abord la nature du sens et son rapport à la référence, puis la délimitation des catégories ainsi que des concepts. Le statut dénominatif des unités lexicales, au cœur de ce débat, devient très important dans la théorie terminologique.

Une question se pose : comment définir la polysémie en terminologie ? Quand existe-t-il polysémie ? Comment l'analyser ? À quel niveau sémantique ou pragmatique ?

Dans la théorie wustérienne des années trente, le terme doit être monosémique, mais nous pouvons observer qu'il se laisse polysémiser ; ce ne sont pas des polysémies de discours, mais des polysémies stables qui rendent compte des nouvelles propriétés du concept scientifique. Ce phénomène peut être étudié d'un point de vue de la discontinuité (sens discrets) ou d'un point de vue de la continuité sémantique (caractère graduel et continu du sens polysémique).

La conception « fixiste de la polysémie fixe en langue un ensemble fini de sens, s'oppose à une conception dynamique ou constructiviste du sens ; dans cette conception, la polysémie par la multiplicité de sens se situe plutôt dans le discours et non pas dans la langue » (Martin 2005). Le constructivisme donne une grande importance au continuum. Entre deux significations ou polysémies il y a toujours des effets de sens intermédiaires.

Quels que soient les problèmes posés par ce phénomène, l'analyse de la polysémie se fait par le « critère référentiel » (Kleiber 1999). La délimitation de la multiplicité des sens est faite à partir des différences d'emploi ou d'usage attestées dans les *corpus* écrits ou oraux ; nous observons que de nouvelles particularités du concept scientifique se trouvent dénotées.

La polysémie est à l'origine d'une variation référentielle. Ce changement n'est pas toujours facile à identifier, mais le changement de référent n'implique pas nécessairement une nouvelle polysémie.

Le contexte linguistique écrit ou oral, en langue de spécialité, joue un rôle dans la sélection de nouveaux sens alors que le contexte situationnel étroit ou large rend possible une délimitation plus efficace de la multiplicité des sens.

À titre indicatif, nous donnons quelques exemples de phénomènes de polysémisation :

[1] *tensão arterial* (portugais) — pression artérielle (français)

1. nous dénotons une restriction de sens observable dans le contexte :

tensão das paredes arteriais correspondentes à pressão que circula o sangue nas arterias (tension des parois artérielles correspondant à la pression du sang qui circule dans les artères) ;

2. nous observons une extension de sens : *pressão sob o qual o sangue circula no interior das artérias* (pression sous laquelle circule le sang dans les artères).

[2] *tumor* (portugais) – tumeur (français)

1. polysémie : benignidade (bénigne)

2. polysémie : malignidade (maligne)

Ces polysémies créent des ambiguïtés dans la communication spécialisée. Cela fait apparaître de nouvelles dénominations : *tumor* (tumeur) devient un terme générique ; deux termes spécifiques ont été créés : *tumor benigno* (tumeur bénigne), *tumor maligno* (tumeur maligne).

Un certain nombre de nouvelles significations ou de nouveaux termes suivent ces principes terminogéniques. Dans le domaine de la médecine générale, nous notons des relations entre des phénomènes de polysémie et de synonymie. Nous présentons les exemples suivants :

- *doença* maladie
- *síndrome* syndrome
- *sistema* système
- *risco* risque
- *traço* trait
- *tensão* tension
- *teste* test
- *prova* examen
- *prova de esforço* test d'effort

En définitive, ces termes polysémiques ne cessent d'acquérir de nouvelles significations et de créer des relations synonymiques avec d'autres termes médicaux. Beaucoup de ces polysémies sont une conséquence des nouvelles particularités cognitives du concept ; d'autres relations de type sémantique s'établissent entre le sous-système de la langue médicale et les sous-systèmes d'autres langues de spécialité. Néanmoins ces phénomènes sont également observables entre les langues de spécialité et la langue générale.

4. En guise de conclusion, le dictionnaire en support papier ou informatisé devient indispensable au traitement de la polysémie ; il présente les polysémies stabilisées et, par la même occasion, il devient un outil utile à l'identification de nouvelles polysémies.

Les dictionnaires électroniques ainsi que les dictionnaires-machine permettent, en général, une reconnaissance plus rapide et efficace des polysémies en langue ; les néologismes sémantiques par polysémisation sont plus difficiles à identifier. Le traitement automatique des langues de spécialité peut contribuer à la détection des termes polysémiques, à partir de *corpus* étiquetés avec l'aide de dictionnaires automatisés, selon des critères sémantiques, lexicaux et syntaxiques.

BIBLIOGRAPHIE

- Cadiot (P.), Habert (B.), 1997 : « Aux sources de la polysémie nominale », dans *Langue Française* 113, pp. 3.
- Condamines (A.), 2005 : *Sémantique et corpus*, Paris, Hermes, Lavoisier, pp. 18.
- Costa (R.), 2001 : *Pressupostos teóricos e metodológicos para a extração automática de unidades terminológicas multilexémicas*, thèse de Doctorat en terminologie, Lisboa, Universidade Nova de Lisboa.
- Cruse (A.), 1986 : *Lexical semantics*, Cambridge, Cambridge University Press.
- Garcia (L.), 2000 : *CD-Rom Dicionário Terminológico de Senologia*, Lisboa, Universidade Nova de Lisboa e Sociedade Portuguesa de Senologia.
- Gross (G.), 1996 : *Les expressions figées*, Paris, Ophrys.
- Martin (R.), 1983 : *Pour une logique du sens*, Paris, PUF.
- Martin (R.), 2005 : « Traitement automatique de la polysémie élogé du dictionnaire », dans *La Polysémie*, Paris, Presses de l'Université Paris-Sorbonne, pp. 168.
- Mejri (S.), sous la dir., 2004 : « Polysémie et polylexicalité », dans *Syntaxe et Sémantique* 5, Caen, Presses Universitaires de Caen.
- Mel'čuk, (I.), Clas (A.), Polguère (A.), 1995 : *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, Duculot, pp. 156-157.
- Kleiber (G.), 1999 : *Problèmes de sémantique. La polysémie en questions*, Lille, Presses du Septentrion.
- Kleiber (G.), 2005 « Quand y a-t-il sens multiple ? Le critère référentiel en question », dans *La Polysémie*, Paris, Presses de l'Université Paris-Sorbonne, pp. 51.
- Rastier (Fr.), 2005 « Enjeux épistémologiques de la linguistique de corpus », dans *La linguistique de corpus*, Rennes, Presses Universitaires de Rennes, pp. 32.
- Victorri (B.), Fuchs (C.), 1996 : *La polysémie. Construction dynamique du sens*, Paris, Hermes, pp. 11.

LA CRÉATIVITÉ LEXICALE DU FRANÇAIS EN RDC : UN CAS D'ANALYSE LEXICOLOGIQUE EN CONTEXTE.

Willy Ilunga Ntumba

département des lettres et civilisation françaises

Université de Kinshasa

Kinshasa, République démocratique du Congo

***Résumé :** La situation politique qui prévaut au Congo-Kinshasa depuis 1990, subsume le travail néologique des Congolais en français. La créativité lexicale se produit ainsi dans un contexte situationnel dominé par les idéologies politiques qui caractérisent le quotidien et les aspirations du peuple. Le processus de démocratisation offre certes aux Congolais l'opportunité d'exprimer autrement leurs pensées mais aussi les nouvelles réalités de la vie qu'ils sont désormais obligés de mener.*

***Mots-clés :** lexicologie, néologie, contexte situationnel, sens des néologismes*

1. INTRODUCTION

Inhérente à la culture et à la nature de l'homme, la créativité lexicale est une activité langagière motivée par la concrétisation de l'ouvrage humain qui plus est reste toujours soumis à la loi de l'évolution. En RDC (République démocratique du Congo), le domaine politique se révèle être le terrain de prédilection où des congolismes se créent davantage.

La présente étude s'évertue certes à analyser en synchronie la néologie congolaise dans le domaine politique et, pour ce faire, elle en a choisi la période de transition, pour des raisons évidentes. En effet, la coutume veut que les époques de grandes transformations politiques et sociales laissent émerger de nouvelles idées qui se traduisent au moyen de nouvelles lexies jusqu'alors inconnues ou peu connues du public. Et la transition entamée depuis le 24 avril 1990, a permis justement une riche créativité lexicale en français.

2. INDICES DE LA CRÉATIVITÉ LEXICALE DANS LA POLITIQUE CONGOLAISE

Pour affirmer qu'il y a créativité lexicale dans l'exercice de l'activité politique au Congo, nous avons procédé préalablement à « l'exploration lexicographique »¹ (Boulanger 1979 : 40). Grâce à la lecture soutenue et intéressée des journaux paraissant à Kinshasa, nous répertorions des mots qui ne sont pas encore attestés dans un dictionnaire ou ouvrage lexicographique de référence. La littérature classique n'étant pas abondante, la presse tend donc à représenter le contenu littéraire du pays, sachant qu'elle constitue en outre le lieu par excel-

¹ « L'exploration lexicographique permet de distinguer les unités reconnues comme lexicalisées, c'est-à-dire consignées dans un dictionnaire, des unités néologiques, c'est-à-dire non encore enregistrées au dictionnaire. »

lence où transparait le plus l'incidence de la donne politique congolaise. Quelques lexies nouvelles proviennent du débat politique radiodiffusé ainsi que du discours officiel et/ou non officiel des politiciens.

Ce répertoire peut être consulté dans un article sous presse intitulé « L'usage du français en RDC : problématique et état des lieux », que nous avons confié au Professeur Ambroise Queffélec pour publication. Nous n'en reprenons pas les contextes syntaxiques parce que ceux-ci sont consignés dans l'article précité et que plausiblement, notre étude s'intéresse plus au contexte situationnel.

Les congolismes sémantiques répertoriés sont : accord global et inclusif, acquis de la conférence nationale souveraine, agression, aigle (aigle de kawele), aile, allié, belligérant, camp de la patrie, combattant (ante), composante, dialoguiste, entité, espace présidentiel, géopolitique, guerre d'agression, guerre de libération, multipartisme intégral, mouvance (ou mouvance présidentielle), parlementaire débout, parti alimentaire, soldat du peuple, transition, transition pacifique et non conflictuelle, union sacrée (ou union sacrée de l'opposition), un plus quatre, etc.

Il existe des congolismes lexématiques comme : afdlien ou afdélien (ienne), aigletesque (ou aigletique), arégionalisation, bembiste, bien mal acquis, congolais rwandophone (ou rwandophone), clientélisé (ée), congolité, déballable, démocrature, démobotisation, démobotiser, détourneur, dialogue intercongolais, dialogueur, ex-faz, facilitateur, gizenguiste, incontournabilité, insécuriser, insectiser, josephiste, journée ville-morte (ou ville-morte), kabiliste, katangalisation, katangaliser, kasaïphobie, kengisme, kenguiste, kibassiste, ministriellement, mouvancier, ngwashiser, pprdien ou péperdien, perdiemiste, préconcertation, prédialogue, prénégociation, primacurien, primature, primaturable, pro-kabiliste, pro-tshisekediste, sectionnaire, taupage (ou tauperie), tshisekedisme, tshisekediste, etc.

Il y a des emprunts locaux : asumba na nganda, banyamulenge (pl. de munyamulenge), bilulu (pl. de kilulu), fwifwi, kadogo, kata-kata, kihuta, maï-maï, Mzee, etc.

Les lexies nouvelles se créent ici par :

1. la dérivation (p. ex. : *déballable*, *intercongolais (aise)*, *ministriellement*, etc.), surtout la suffixation des noms propres (p. ex. : *bembiste*, *josephiste*, *tshisekediste*, etc.), des sigles des partis (p. ex. : *afdélien*, *mprien*, *péperdien*, *udépésien*, etc.) ;
2. la composition (p. ex. : *Congolais rwandophone*, *dialogue intercongolais*, *espace présidentiel*, *guerre d'agression*, etc.).

Elles se forment aussi à partir de l'existence des mots virtuels (p. ex. : *insécuriser*, *incontournabilité*, *démocrature*), des emprunts aux langues étrangères, p. ex. : *perdiemiste* (de *per diem*), *primacurien* (de *prima curia*), des emprunts locaux (p. ex. : *asumba na nganda*, *kadogo*, *kata-kata*, *maï-maï*, etc.).

3. CONTEXTE DE FORMATION DU VOCABULAIRE POLITIQUE CONGOLAIS

La transition congolaise commence le 24 avril 1990. Le discours pathétique qui l'annonce reprend pour la première fois dans le vocabulaire du président Mobutu des mots nouveaux tels que : *démocratie, démocratisation, multipartisme, parti politique*, etc. Depuis, la RDC connaît une situation politique sur fond de crise, « récit d'une constante frustration, d'un éternel blocage... » (Peter Rosenblum (postf.) dans Verhaegen (1996 : 502)).

Consécutivement, les Congolais n'ont pas hésité à recourir à la néologie afin de s'accommoder au quotidien de la vie, aux impératifs de la réalité qui se présente à eux et à laquelle ils se confrontent. Chaque terme créé dénote le besoin de dire ou de nommer autrement, et la créativité lexicale devient ainsi le symbole de « la liberté d'invention de l'individu – ou du groupe – face à l'oppression des pouvoirs constitués » (Bastuji 1979 : 12).

L'intérêt politique à travers la presse locale provoque notamment le phénomène *parlementaire débout* « personne qui discute et commente la politique à longueur des journées », s'agglutinant devant les kiosques ou les étalages des journaux. Et comptés parmi les *combattants* « purs et durs » de l'*opposition radicale*, les *parlementaires débout* jouent en effet un rôle important dans la diffusion des mots nouveaux.

Des faits et contrefaits successifs s'expriment par une terminologie particulièrement contextualisée, suivant les temps forts ci-après :

- période avant la conférence nationale souveraine (1990-1991),
- période pendant la conférence nationale souveraine (1991-1992),
- période après la conférence nationale souveraine (1992-1997),
- période avant le dialogue intercongolais (1997-2001) et
- période après le dialogue intercongolais (2001 à nos jours).

3.1. Le vocabulaire d'avant la Conférence nationale souveraine

Le début de la transition est marqué visiblement par un antagonisme dialectique et discursif, signe évident d'un espace politique bicéphale où le pouvoir et l'opposition – qui s'organise en l'*Union sacrée*, coalition des *forces du changement* – s'affrontent même au niveau du travail néologique. *Pouvoir* et *opposition* se généralisent à cette époque qui voit la dictature côtoyer un *alter ego*, la démocratie.

La cohabitation de ces deux concepts n'est pas sans accros. Le pouvoir se prévaut d'avoir initié le processus de démocratisation, donc d'avoir instauré la démocratie au pays. L'opposition ne le lui reconnaît nullement, d'autant que le dictateur n'arrive pas à démocratiser son régime. Le génie de la néologie crée *démocrature* « la démocratie du dictateur ». *Dictature* égal pouvoir de Mobutu, *démocratie*, pouvoir du peuple, et *démocrature*, système permettant au dictateur d'adapter sa politique à la nouvelle réalité.

Les Congolais découvrent ensuite des lexies comme *multipartisme à trois* et *multipartisme intégral*. Le « multipartisme à trois » voulu par Mobutu, ouvrait l'espace politique aux activités de trois partis. Cette limitation est vite récusée par l'opinion politique appelant au « multipartisme intégral » qui a donné accès à la prolifération des partis.

La lutte pour la démocratisation véritable du Congo-Kinshasa plonge progressivement le pays dans une crise sociale, politique et économique aux conséquences incommensurables. Le peuple fonde son espoir sur le combat de l'opposition, caractérisé par la volonté du *changement* : le système de gestion calamiteuse de Mobutu devait disparaître en vue de l'émergence d'un nouveau système de gestion, fondé sur des valeurs démocratiques. Ainsi, *changement* s'oppose au *statu quo* « le fait, pour les *mobutistes* (« partisans de Mobutu »), de vouloir garder et pérenniser le système ancien ».

Mobutu peut bien dire, de façon expressive : « Tout doit changer, tout va changer », mais l'on doit le *changement* à Tshisekedi qui, à certains égards, incarne ce combat politique. Sa famille politique se nomme « *forces du changement* » et ses adeptes, partisans et sympathisants (*tshisekedistes*) s'appellent « *combattants* ».

Au demeurant, la reconversion des *mobutistes* semble difficile. On les retrouve dans l'opposition, beaucoup plus au sein de l'UDI (Union des démocrates indépendants), mais aussi ailleurs. L'UDI est surnommé « Union des détourneurs impénitents » et/ou « Union des dinosaures impunis ». *Détourneur* et *dinosaure* désignent « tout collaborateur de Mobutu s'étant enrichi au détriment du peuple ». Dans l'opposition ou au pouvoir, ceux qui ont cette marque sont appelés *mouvanciers*. À la suite de Tshisekedi, on les appellera *caciques du MPR*. Ils sont tous *primacuriens* « du *prima curia* »².

3.2. Le vocabulaire de la CNS

Le souci de résorber la crise congolaise conduit la volonté collective à organiser la CNS (Conférence nationale souveraine), lieu où l'État doit être exorcisé et le « *mal zairois* » – signifiant Mobutu et sa façon de gouverner – extirpé, et enfin, baliser les pistes de la *réconciliation nationale*. Avec la CNS, les Congolais s'habituent à l'écoute et à l'usage des expressions telles que : *abstention*, *adopté*, *consensus*, *motion*, *per diem*, *plénière*, *quorum*, *vote à main levée*, etc. comme s'il n'y avait jamais eu d'institution parlementaire au pays.

Face à la précarité et la déliquescence de son pouvoir, Mobutu monte un stratagème machiavélique pour se maintenir. Il use du débauchage (au sein des partis, voire des *plateformes*), et provoque le *vagabondage politique*, mais il est contraint de négocier avec l'opposition. La première rencontre officielle entre le pouvoir et l'opposition au « *palais de marbre* » a lieu en 1991 et d'autres s'ensuivront. Naissent ainsi les paradigmes *négociations*, *consultations*, *concertation (s) débat* et *dialogue* ; chacun de ces termes étant employé dans une situation particulière. Avec l'objectif de raviver la nation, ces tractations sont sanctionnées par la signature d'un accord. D'où les paradigmes *Compromis politique*

² Société secrète (créée en 1985) dont l'existence a été révélée (en 1990) par un journaliste, réunissant les dignitaires mobutistes avec la mission de pérenniser le MPR quel qu'en soit le prix. Et selon une considération populaire, astreints aux sacrifices malfaisants, les *primacuriens* sont auteurs de la débâcle du pays.

global (1992), *Protocole d'accord* (1994), *Accord de Lusaka* (1999), *accord de Sun City* (2002), *accord global et inclusif* (2002), etc.

Les rencontres des politiciens aboutissent vraisemblablement à la mise en place d'un nouveau gouvernement, à l'instauration d'un « nouvel ordre politique ». Chaque gouvernement porte un nom : *gouvernement de transition* (1990, 1992, 1994, 2003), *gouvernement de transition élargi* (1990), *gouvernement de combat* (1991), *gouvernement de large union nationale* (1991), *gouvernement de salut public* (1991, 1997), *gouvernement d'union nationale* (1992), *gouvernement de large union nationale et de salut public* (1993), *gouvernement de salut national* (1996). Ces gouvernements se forment sur base d'un nouveau critère, la *géopolitique*, c'est-à-dire « respect de l'équilibre ethnique et géographique ».

La recrudescence de l'insécurité à Kinshasa, délibérément voulue par le pouvoir pendant cette période, sert à *insécuriser*, c'est-à-dire « mettre en insécurité » les opposants. Chargés de dissuader les « confrenciers » de déballer davantage le système de Mobutu, les *hiboux*³ entrent en transe. Mais ce système reste *déballable*, « qui peut être déballe » à tout point de vue. C'est d'ailleurs après ce déballage qu'on a pu identifier les détenteurs des *biens mal acquis*.

Aux côtés des hiboux, *asumba na nganda* « personnage mythique, invisible et invincible, qui se distingue par des vols et assassinats surtout commandités », fait son apparition. Il deviendra l'éponyme des personnes qui s'adonnent à cette sale besogne, mais aussi de celles qui excellent dans les abus.

3.3. Le vocabulaire d'après la CNS

La clôture en catastrophe de la CNS impose le respect des *acquis de la conférence nationale souveraine*, synonyme de maintien de Tshisekedi (premier ministre élu à la CNS) à la *primature* « poste de premier ministre, son bureau ». Tous les autres *primaturables* (personnes à même de devenir premiers ministres) sont donc nommés illégalement. On clame son *incontournabilité* « caractère de celui que l'on ne peut contourner ». L'opposition radicale focalise son combat politique sur le respect de ces acquis. Elle organise des marches pacifiques, grèves, *journées ville-morte* « paralysie volontaire et collective des activités socioprofessionnelles à travers la ville ».

Le comportement « indigne » des politiciens inspire des métaphores assez significatives, plus par l'animalisation. Mobutu, le « vieux léopard » (toujours prêt à dévorer ses adversaires), devient *aigle* puis *aigle de kawele* quand il se dit : « Je suis le chapiteau » ou «... au-dessus de la mêlée » sans l'être véritablement. Car il débauche en corrompant pour se constituer la *majorité présidentielle*. Avec aigle se forment les adjectifs *aigletesque* et/ou *aigletique* (« aquilin ») pour qualifier ce comportement.

Le débauchage résulte du *vagabondage politique*, œuvre des *taupes* qui infiltrent l'opposition radicale et font preuve de *tauperie* ou de *taupage* « infiltration, espionnage ».

³ Nom donné aux spadassins de Mobutu (soldats entraînés aux techniques non conventionnelles : guérilla urbaine, sabotage, enlèvement, répression, etc.). À la manière de l'oiseau du même nom, les hiboux opèrent surtout la nuit.

Mungul Diaka, l'une de ces taupes, prétend être « l'homme des situations difficiles » et s'appelle *kihuta* (du kikongo « vipère ») dont le venin neutraliserait la radicalisation de certains opposants.

Nguz a Karl-i-Bond cautionne la *kasaïphobie* « hostilité aux Kasaiens » au Katanga. Pour lui, les Kasaiens sont des *bilulu* (du swahili « insectes ») « vermines », mais il ne fait pas exception à cette qualification et paraît même plus nuisible, donc *kilulu premier*. Il est *fwifwi* (du swahili « hibou ») à cause des grosses lunettes qui surplombent ses yeux.

Une profonde crise institutionnelle se déclare ensuite. Celle-ci nécessite une certaine médiation, mais monseigneur Monsengwo (*médiateur*) et la *Troïka* (triple France, Belgique, États-Unis) n'y apportent aucun dénouement.

La guerre de l'Est – *guerre de libération* pour les « *kabilistes* » et *guerre d'occupation* pour le pouvoir en place – (octobre 1996), commence et pousse Mobutu à la fuite (16 mai 1997).

La guerre de l'Alliance des forces démocratiques pour la libération du Congo-Kinshasa dont les initiales servent à former le dérivé *afdélien* « de l'AFDL », est une occasion pour la population du Kivu (*munyamulenge*), rwandaise de souche, de revendiquer la nationalité congolaise. Et les *Banyamulenge* (du swahili, habitants de Mulenge) n'existent pas selon une certaine opinion, car les ethnies congolaises portent généralement le nom de leur langue (p. ex. : *Baluba*, locuteurs du ciluba, *Bayaka*, locuteurs du yaka, etc.).

3.4. Le vocabulaire prédialogue intercongolais

La déclaration de prise de pouvoir (17 mai 1997) de Laurent-Désiré Kabila, affectueusement appelé *Mzee* (du swahili « le chef »), déclenche une révolution terminologique. Le pays est rebaptisé et avec lui la toponymie imposée par le « recours à l'authenticité » est assez bousculée, ainsi que la nomenclature du découpage territorial.

Ne pouvant pas se contenter d'une armée composée de combattants étrangers, Kabila recrute des *kadogo* (du swahili, « petit »), qui se distinguent par leur petite taille et leur âge très jeune.

La victoire de l'armée de l'AFDL est marquée par la mise en déroute des éléments des FAZ (Forces armées zaïroises), soumis à l'exil ou à l'intégration des Forces armées congolaises. On les appelle *ex-faz*, et leur comportement est tout à fait antagonique face aux *kadogo* dont ils méprisent la formation.

Pour avoir gagné la guerre de libération, Kabila devient *libérateur* « qui a libéré le peuple congolais de l'emprise du dictateur », mais son pouvoir ne va pas tarder à être taxé de dictature. On lui reproche notamment la *katangalisation* (« action de placer des Katangais ») des secteurs clés de la vie nationale.

Certains ex-opposants de Mobutu rejoignent les rangs de l'AFDL, d'autres se muent en opposants de Kabila. Le mécontentement qui se déclenche au sein de la famille *afdélienne*, provoque la « *guerre d'agression* » (2 août 1998). Afin de résister aux *agresseurs*, Kabila appelle au secours ses compères de l'Angola, de la Namibie et du Zimbabwe, les *alliés*.

La « guerre d'agression » fait l'objet de plusieurs controverses : le pouvoir met en exergue une agression et les mécontents de l'AFDL revendiquent une rébellion. La dichotomie pouvoir/opposition se ramifie ainsi pour donner lieu à un espace politique tricéphale (gouvernement de Kinshasa, opposition armée et opposition politique). Cette trilogie influe sur l'activité néologique à telle enseigne que les lexies créées reflètent naturellement leur appartenance à l'une ou l'autre de ces composantes.

Agression a un sens restreint : elle se définit comme l'invasion du territoire congolais par les armées du Rwanda, de l'Ouganda et du Burundi (*agression rwando-ougando-burundaise*).

Dans ce contexte, *agression* est antonyme de *libération* « lutte contre la dictature » autrefois considérée comme une rébellion par le pouvoir de Mobutu.

Comprise à son sens propre par les non-belligérants, *rébellion* représente cependant une mission noble et légitime visant le bien-être collectif aux yeux des tenants de la libération.

De ce fait, *rébellion* et *libération* se confondent dans leur contenu sémantique, car elles permettent au peuple de devenir « libre » après l'éviction du régime dictatorial combattu.

Sur base des considérations idéologiques, ces unités lexicales connotées entretiennent donc tantôt une relation sémantique de synonymie tantôt d'antonymie.

Pour l'opposition politique, *rébellion*, *guerre d'agression*, *guerre de libération* sont synonymes ; elles renvoient à une « guerre de multinationales » qui a dévasté le pays, ses ressources et sa population. C'est le cas avec *rebelle*, *libérateur*, *agresseur*.

Pour les belligérants, *rébellion*, *guerre d'agression*, *guerre de libération* sont antonymes ; elles répondent aux acceptions mentionnées ci-dessous.

3.5. Le vocabulaire post-dialogue intercongolais

En vue de mettre fin à un conflit dévastateur, la communauté internationale convainc les protagonistes à signer l'accord de cessez-le-feu de Lusaka qui préconise l'organisation du *dialogue intercongolais*.

Le forum devant réunir les Congolais, nécessite un *facilitateur* « qui rend facile ». Dans son rôle de conciliateur, le « facilitateur » recrute les *dialogueurs* ou *dialoguistes* « participants au dialogue intercongolais » d'après leur *composante* « gouvernement de Kinshasa, opposition armée (RCD et MLC), opposition politique et/ou société civile » ou *entité* « faction armée dissidente du RCD ou du MLC, et/ou maï-maï ».

Ayant désigné un vice-président de la République au gouvernement (sauf la société civile), une « composante » est numériquement plus importante qu'une « entité ».

Après la fustigation de l'*accord de Sun City* signé par le *camp de la patrie* (gouvernement de Kinshasa, MLC et leurs alliés), le dialogue intercongolais aboutit à la signature de l'*Accord global et inclusif*.

L'*un plus quatre (1+4)* « système politique fonctionnant avec un président de la République, chef du gouvernement, secondé par quatre vice-présidents pilotant chacun une com-

mission gouvernementale », est instauré. Le président et ses adjoints forment l'*espace présidentiel* « présidence de la République » en tant qu'institution.

Le dialogue intercongolais rétablit apparemment la paix ; ce qui suppose la fin de l'état belligérant. Mais *belligérant* reste actuel et désigne les composantes qui se sont fait la guerre (l'ex-gouvernement représenté par la famille politique de Joseph Kabila, le RCD et le MLC).

La famille politique de Kabila fils éclate en deux tendances : les *kabilistes* qui incarnent prétendument l'idéologie de Kabila père et les *josephistes* qui se reconnaissent du fils. Ils militent au sein du PPRD (Parti du peuple pour la reconstruction et le développement). *Péperdien* (*enne*) en est le dérivé.

Le RCD (Rassemblement des Congolais pour la démocratie) est à majorité tutsie. Il s'agit des Tutsis congolais appelés *rwandophones* (ou *Congolais rwandophones*) à cause de la langue du Rwanda qu'ils parlent.

Parrainé en tant que rébellion par le gouvernement de l'Ouganda, le MLC (Mouvement de libération du Congo) est présidé par Jean-Pierre Bemba dont les adeptes, partisans et sympathisants s'appellent *bembistes*.

L'opposition politique « radicale » prononce l'arrêt de la transition au 30 juin 2005 à la suite du non-respect de certaines clauses de l'*Accord global et inclusif*. Un peu comme pour décourager les manifestants potentiels, le phénomène *asumba na nganda* réapparaîtra dans ce contexte sous le nom de *kata-kata* (du lingala, *kokata* « couper »)⁴ (4).

4. PERTINENCE DU CONTEXTE EN LEXICOLOGIE

Nous considérons la pertinence du contexte en amont comme en aval. En effet, le contexte détermine les conditions de création des mots autant qu'il en détermine le mode d'utilisation ainsi que les postulats de leur compréhension. Nous relevons donc une relation de subsumption réciproque entre la néologie ainsi que l'emploi des néologismes, d'un côté, et le contexte, de l'autre.

Les corollaires de la démocratie (liberté de presse, liberté d'opinion et d'expression...) font de la transition une période propice à l'activité de création et de découverte de nouveaux mots.

Érigé en une dictature prohibant toute politique à l'encontre des idéaux du MPR Parti-État, le régime de Mobutu n'a pas permis au peuple, du reste subjugué et muselé, de s'adonner à la créativité et produire des mots de notre *corpus*. Le discours présidentiel du 24 avril 1990 apparaît donc comme un déclic qui a délié la langue du peuple.

Le contexte joue aussi un rôle important dans la conception des sèmes d'un terme linguistique. En effet, la définition du contenu sémantique se fait en contexte, et hors contexte sa

⁴ Du lingala, *kokatakata* signifie découper. Ainsi donc, *kata-kata* c'est un assassin (boucher) qui, après son meurtre, coupe le corps de la victime en morceaux.

valeur est *ipso facto* appréhendée différemment par les membres d'une même communauté linguistique, voire par les actants d'une même interlocution.

Par exemple, Laurent-Désiré Kabila se considère « rebelle » pendant qu'il est président de la République, qualificatif qu'il a refusé obstinément de s'attribuer quand il faisait la guerre au pouvoir de Mobutu et même d'attribuer à ceux qui lui ont fait la guerre à partir du 2 août 1998. *Rebelle* se comprend ici par rapport à une acception particulière : « un révolté, peu importe son statut, toujours prêt à contester tout ce qui ne cadre pas avec sa vision ».

Cette référence se précise par le fait que la lutte armée de Kabila n'est pas conçue par les *kabilistes* comme une rébellion d'autant plus que le peuple, seule source de légitimité du pouvoir, l'a soutenue. Pour comprendre cette particularité, il faut faire appel aux propriétés afférentes du mot, lesquelles propriétés constituent des « traits sémantiques dont l'actualisation résulte d'une contrainte contextuelle » (Rastier *et alii* 1994 : 38).

Kabila n'est donc pas rebelle ; le pouvoir de Mobutu contre lequel il s'est battu n'en était d'ailleurs pas un parce que désavoué par le peuple. Il est le *soldat du peuple* et pour ce mérite, il est plébiscité *héros national* après son assassinat.

S'il doit se définir comme l'ensemble des paramètres significatifs ayant une incidence sur une unité sémantique, et *vice versa*, le contexte auquel nous faisons allusion est un contexte situationnel, cadre dans lequel l'énoncé est produit.

Le contexte situationnel diffère du contexte syntagmatique qui permet d'attribuer le sens aux homonymes pour ainsi arriver à exclure entre eux la relation de synonymie qui les confondrait. Il s'oppose par ailleurs au contexte paradigmatique qui permet, lui, de nuancer le sens des synonymes. Il peut être considéré comme le moyen « extralinguistique » de limitation de l'extension d'une lexie (Picoche 1977 : 64) et s'avère tout aussi important que le contexte syntaxique qui en constitue le moyen typiquement linguistique et auquel il reste étroitement lié.

Le contexte d'emploi d'une unité lexicale permet certes d'atteindre l'actualisation qui est l'opération interprétative à même d'identifier ou de construire un sème, et le contexte situationnel doit être identifié préalablement aux fins de circonscrire l'énoncé, de circonscrire et d'appréhender la pensée. Aussi considérons-nous la compétence communicative comme la faculté de choisir des formes linguistiques dans une situation particulière et de leur attribuer non seulement un sens référentiel, dénotatif, mais également le sens social et affectif.

5. CONCLUSION

La créativité lexicale du français en RDC telle que présentée dans cette étude, démontre en définitive que les contraintes situationnelles aident à reconstituer le sens exact d'une unité lexicale polysémique selon le contexte de sa création et/ou de son emploi.

La crise politique qui sévit au Congo-Kinshasa depuis le lancement du processus de la démocratisation, inspire un fonds terminologique particularisant. Ce fonds se constitue grâce à la liberté de l'expression et de la presse, apparente par rapport à la prolifération des titres des journaux, des chaînes de radiodiffusion et de télévision, mais aussi et surtout à

l'exercice des activités politiques où le *feed-back* entre leaders et militants des partis n'est jamais interrompu.

À la lumière des renseignements nous légués par Louis Guilbert (1975 : 40-44), le travail néologique dans le contexte politique de la RDC, répond à la nécessité de nommer un objet ou un concept nouveau, une réalité nouvelle (« néologie dénominative »). Il s'ajoute à cela la volonté de dire expressivement ce que l'on pense et de manière inédite (« néologie stylistique »). En outre, l'existence des mots virtuels suscite la création des lexies nouvelles qui répondent parfaitement à la norme morphosyntaxique du français (« néologie de langue »). Et comme le souligne André-Marcel d'Ans, « l'humour et la dérision jouent souvent un rôle réjouissant dans cette production lexicale pittoresque » (1995 : 12).

L'analyse lexicologique du vocabulaire politique de la transition congolaise nous permet d'affirmer d'une part qu'en amont, le contexte situationnel représente les conditions d'émergence des néologismes. En aval, ce contexte détermine la valeur sémantique réelle des termes linguistiques, que l'on doit établir autant à leur emploi qu'à leur compréhension dans un contexte syntaxique.

BIBLIOGRAPHIE

- Adda (R.) *et alii*, 1979 : *Néologie et Lexicologie*, Paris, Larousse « Langue et langage ».
- Ans (d') (A.-M.), édit., 1995 : *Langage et Politique. Les Mots de la démocratie dans les pays du sud de l'espace francophone*, Paris, CIRELFA-ACCT/Didier Érudition « Langue et développement ».
- Bastuji (J.), 1979 : « Notes sur la créativité lexicale », dans Adda (R.) *et alii*, *Néologie et Lexicologie*, Paris, Larousse, pp. 12-20.
- Boulangier (J.-C.), 1979 : « Problématique d'une méthodologie d'identification des néologismes en terminologie », dans Adda (R.) *et alii*, *Néologie et Lexicologie*, Paris, Larousse, pp. 36-44.
- Cusin-Berche (F.), 2003 : *Les Mots et leurs contextes*, Paris, Presses Sorbonne nouvelle.
- Depecker (L.), 1990 : *Les mots de la Francophonie*, Paris, Belin.
- Edema Atibakwa (B.), 1998 : *Étude lexico-sémantique des particularismes français du Zaïre*, thèse de doctorat, Aix-en-Provence, inédit.
- Équipe IFA, 2004 : *Inventaire des particularités lexicales du français en Afrique noire*, Paris, EDICEF-AUF.
- Gayral (F.), 1998 : « Hypothèse de compositionnalité et créativité du sens en contexte », dans *Traitement automatique des langues*, vol. XXXIX, n° 1, pp. 67-98.
- Guilbert (L.), 1975 : *La Créativité lexicale*, Paris, Larousse « Langue et langage ».
- Ilunga (N.), 2005 : *L'Usage du français en RDC : problématique et état des lieux*, inédit.
- Latin (D.), Queffélec (A.) et Tabi-Manga (J.), édit., 1993 : *Inventaire des usages de la francophonie : nomenclatures et méthodologies*, Paris, AUPELF-UREF/John Libbey Eurotext « Actualité scientifique ».
- Lesay (J.-D.), 2004 : *Les Personnages devenus mots*, Paris, Belin.
- Picoche (J.), 1977 : *Précis de lexicologie française*, Paris, Nathan.
- Pruvost (J.) et Sablayrolles (J.-F.), 2003 : *Les Néologismes*, Paris, PUF « Que sais-je ? ».
- Rustier (F.), Cavazza (M.) et Abeillé (A.), 1994 : *Sémantique pour l'analyse. De la linguistique à l'informatique*, Paris/Milan/Barcelone, Masson « Sciences cognitives ».
- Verhaegen (B.) (dir.), 1996 : *Zaïre 1992-1996. Chronique d'une transition inachevée*, Paris, L'Harmattan.

LA PRESSE VUE À TRAVERS *NÉOSCOPE* : QUAND LES CONTEXTES MÉDIATIQUES SONT MIS AU SERVICE DE LA NÉOLOGIE

Lina Sader Feghali

Centre d'études et de recherches en terminologie arabe (CERTA)

École de traducteurs et d'interprètes de Beyrouth (ETIB)

Université Saint-Joseph, Beyrouth, Liban

lina.feghali@usj.edu.lb

Résumé : *Cette communication a pour objectif de présenter le projet de recherche entrepris au Centre d'études et de recherches en terminologie arabe (CERTA) qui vise à recenser les néologismes parus dans la presse. L'examen des problèmes rencontrés et des solutions envisagées lors du repérage des néologismes et de l'analyse du corpus met en question le statut des néologismes médiatiques retenus puisqu'ils sont issus de contextes journalistiques, une denrée périssable où l'éphémère, l'à-peu-près et le sensationnel règnent.*

Mots-clés : *néologismes, contextes médiatiques, base de données terminologiques, paramètres d'inclusion et d'exclusion*

Dans ce monde en pleine mutation, nul ne peut nier l'importance de l'actualité pour être au courant et à la page. La consultation des médias est en fait le moyen par excellence pour d'abord acquérir des informations, voire des connaissances – dans presque tous les domaines – et aussi pour se divertir et prouver son appartenance à la société. Elle permet aussi d'avoir accès à un lexique, sans cesse remis à jour, de mots de passe qui servent au décryptage de l'information et à sa compréhension. Les médias sont enfin le bassin privilégié pour la diffusion des créations néologiques qui témoignent de l'apparition d'une nouveauté ou qui répondent à un besoin d'expression et de communication.

C'est ce dernier aspect spécifique du discours médiatique qui a fait l'objet d'un projet de recherche entrepris par le Centre d'études et de recherches en terminologie arabe (CERTA) de l'École de traducteurs et d'interprètes de Beyrouth (ETIB). Ce projet avait pour objectif de créer une cellule de veille qui repère les néologismes médiatiques, en vue de les soumettre à une analyse terminologique et de les saisir dans *Néoscope*, une base de données terminologiques en ligne. Cette communication passe en revue les enjeux d'un tel projet car l'analyse contextuelle, principalement basée sur des contextes médiatiques, met en question le statut des néologismes : sont-ils des néologismes au vrai sens du terme, dans quelle mesure un *corpus* médiatique peut-il être fiable ? Et, par conséquent, peut-on qualifier de néologismes des dénominations qui ne durent parfois que le temps nécessaire pour que l'encre sèche ?

Après la présentation du projet en tant que tel, dans une première partie, nous nous pencherons dans une deuxième partie sur les problèmes rencontrés et les solutions envisagées lors du repérage des néologismes et de l'analyse du *corpus*.

1. PRÉSENTATION DU PROJET

Le projet CERTA *en ligne* a été lancé en juillet 2004 dans le cadre d'un contrat d'association JER 1015 conclu avec l'AUF. Ce projet a prévu la mise en place d'un site qui offre aux usagers trois services différents, à savoir :

- repérer, consigner et diffuser les néologismes parus dans les médias dans *Néoscope* ;
- répertorier les sites à intérêt terminologique dans *Cyberterminologie* ;
- et profiter d'un service d'assistance terminologique grâce à *SOS-Termes*.

La base *Néoscope*¹, qui est l'objet de cette communication, est née du désir de proposer une base de données terminologiques² qui se caractérise par son originalité et sa spécificité. Elle est originale de par son contenu, non seulement parce qu'elle n'est pas identique à celui des autres supports terminologiques existants, mais aussi parce que le *corpus* qui a servi de base au travail de recherche est essentiellement constitué de contextes médiatiques tirés de quotidiens en langue arabe et française³, donc qui touchent à un domaine essentiellement dominé par l'éphémère, l'à-peu-près et le sensationnel. Pourquoi le choix de ce domaine spécifique qui ne répond pas aux conditions d'un domaine de spécialité en bonne et due forme ? Trois raisons expliquent ce choix, à savoir : le court délai d'exécution du projet (juin 2004 – janvier 2005) qui ne permettait pas la prévision d'une étape de familiarisation avec un domaine de pointe, l'accessibilité des sources et la possibilité de les cerner en vue d'un dépouillement efficace et, enfin, la variété des notions qui ne se limitent pas à un domaine précis. En effet, les quotidiens charrient dans leurs pages un flux journalier de nouvelles en provenance des horizons les plus divers et, de par leur contenu en mosaïque formé de textes hétéroclites, ils abordent des sujets diversifiés qui touchent à différents domaines et, par conséquent, constituent un moyen privilégié de diffusion de nouvelles entités.

La cellule de veille, constituée d'étudiantes en master à l'ETIB, avait pour mission de traquer tout néologisme paru dans les médias, de rechercher le maximum de contextes possibles pour procéder à leur analyse, et de consigner enfin les données dans une fiche diffusée directement en ligne. Quant à la structure de la base de données *Néoscope*, elle a été établie selon un schéma particulier qui s'adapte aux besoins de consignation. Les domaines sont aussi divers que les thèmes d'actualité (actualités politiques, sociales, sportives, technologiques, économiques, militaires, informatiques, spatiales, etc.) alors que les sous-domaines

¹ L'appellation *Néoscope* est forgée avec les formants grecs *néo* (nouveau) et *scope* (observer).

² La base de données a été conçue par le chef du projet, Lina Sader Feghali, et a été développée avec le langage PHP par Nathalie Bitar, Webmestre de l'AUF-Bureau Moyen-Orient. La base *Néoscope*, hébergée sur un serveur Linux du Bureau Moyen-Orient de l'AUF (www.lb.refer.org/neoscope/projet.php), comporte 150 fiches environ.

³ Les quotidiens choisis ont été principalement des quotidiens libanais de grande diffusion dans le monde arabe (*L'Orient-Le Jour*, *An Nahar*, *Al-Hayat*, etc.)

déterminent, s'il y a lieu, le pays directement concerné par la nouvelle réalité. Le bloc *terminologie* de chaque fiche comporte pour les deux langues de travail trois principaux champs : d'abord, le champ propre à la *dénomination* (terme vedette, variantes, synonymes), ensuite, le champ *définition* qui comporte les principaux caractères notionnels tirés à partir des contextes définitoires et explicatifs, et, enfin, le champ *notes* qui contient soit une note d'usage linguistique soit une note qui fournit des informations complémentaires sur la notion elle-même ou sur les circonstances qui ont abouti à son apparition. Certaines fiches sont, en outre, dotées d'une illustration qui éclaire davantage la notion. En plus du bloc relatif aux données terminologiques, un bloc *sources* permet à l'utilisateur de consulter les références des sources utilisées et aussi, dans certains cas, d'avoir directement accès, grâce à des hyperliens, aux références disponibles en ligne. Quant à la consultation de *Néoscope*, elle se fait soit en balayant la liste des entrées classées par ordre alphabétique, soit en utilisant la fonction « Rechercher par » qui permet de rechercher une fiche par terme, par domaine ou par terme et domaine à la fois en français et en arabe.

En consultant *Néoscope*, l'utilisateur découvre des dénominations qui ont marqué l'actualité entre juin 2004 et janvier 2005. Cependant, la consignation de ces dénominations dans *Néoscope* n'est pas pour autant une attestation de leur implantation éventuelle dans l'usage, car certaines sont vouées à disparaître à jamais, alors que d'autres occuperaient une place dans les manuels d'histoire au moment où certaines ont des chances de continuer à marquer notre quotidien. Cet état de fait nous mène directement au cœur du problème, à savoir passer en revue les problèmes et les solutions envisagées à deux niveaux : le repérage des néologismes et l'analyse du *corpus*.

2. PROBLÈMES RENCONTRÉS ET SOLUTIONS ENVISAGÉES

Le fait de choisir l'actualité pour traquer les néologismes est une arme à double tranchant. D'une part, l'actualité est, par définition, l'« ensemble des événements actuels, des faits tout récents » (Robert 1993) ; elle est donc un réceptacle ouvert à toute nouveauté. D'autre part, et de par son essence, l'actualité se défraîchit très vite ; donc, toute nouveauté est exposée au risque de disparaître, sinon de subir des modifications. Ceci signifie-t-il pour autant que les informations transmises dans les quotidiens donnent accès à toutes les nouveautés sans exception ? Ce n'est pas le cas car une nouvelle n'a droit à l'existence que si elle est insolite, inattendue, exceptionnelle ou sensationnelle. Un fait qui a lieu n'est pas nécessairement un événement. Pour qu'il le devienne, il faut que les médias en parlent⁴ et que les thèmes abordés répondent aux attentes des lecteurs. Par conséquent, toute recherche dans ce domaine est tributaire du *corpus* utilisé qui détermine la nature des néologismes et des données terminologiques qui les définissent.

⁴ « L'informatif, c'est ce qui est nouveau dans un message, ce qui ne fait pas partie intégrante et immuable de l'environnement du récepteur [...] Ce qui fait d'un événement une nouvelle, c'est son caractère surprenant, inhabituel (mais inhabituel ne veut pas dire inattendu) (Kientz 1971 : 83) ».

2.1. Repérage des néologismes

Le néologisme est une notion mouvante qui, selon l'angle sous lequel il est considéré, revêt un caractère différent. En règle générale, le statut de néologisme est reconnu à toute innovation lexicale susceptible de devenir une unité lexicale. Cependant, dans le cadre de ce projet, a été considérée comme néologisme toute nouvelle dénomination désignant une nouvelle réalité qui a marqué l'actualité sans pour autant qu'elle soit nécessairement susceptible de devenir une unité lexicale. En effet, est-il possible de refuser aux innovations qui occupent la une de l'actualité et qui deviennent, même pour un moment, le centre des débats politiques, sociaux, technologiques, le statut de néologisme⁵? En guise de réponse à cette question et pour justifier le droit au statut de néologisme à toute innovation qui désigne une nouvelle réalité, une série de paramètres d'inclusion et d'exclusion a été mise à l'épreuve. Des quatre paramètres identifiés par Cabré (1998 : 254) nous avons commencé par celui qui est le plus souvent pris en compte par les néologues, à savoir le paramètre lexicographique qui permet de déterminer si une unité lexicale est néologique en vérifiant sa présence ou son absence dans un *corpus* de dictionnaires prédéterminé. Même si ce paramètre est relativement facilement vérifiable dans ce cas, l'analyse *a posteriori* des unités retenues⁶ a permis de les classer en quatre catégories bien distinctes, et ce en fonction de leur chance d'intégrer un dictionnaire : la première catégorie comprend les unités susceptibles d'intégrer le lexique général (*alterconsommateur*, *chav*⁷, *masstige*, *métrosexuel*, *bioart*, *cybercriminalité*) alors que la seconde comprend les unités susceptibles d'intégrer le lexique spécialisé (*bio-sentinelles* dans le domaine de la biologie, *ascenseur spatial* dans le domaine spatial, *homme de Florès* dans le domaine de la paléontologie, *phishing* ou *hameçonnage* et *blogue* ou *bloc* dans le domaine informatique, *PSC-Rantes* et *Syndrome X* dans le domaine médical, etc.). Quant aux unités faisant partie des deux dernières catégories, elles ne sont pas susceptibles de devenir des unités lexicales même si elles désignent des notions bien précises. En effet, la troisième catégorie comprend les unités propres à un groupe social limité qui font l'objet soit d'un usage restreint (*États indécis*, *toux du WTC* et *milices amies* aux États-Unis, les *hawassem* en Irak, le *Mur des propositions* du Forum social mondial de Porto Alegre, les *reductionnistes* au Liban), soit d'un usage plus répandu à cause de l'intérêt accordé par les médias à une région particulière du monde (*zone verte* et *triangle sunnite* en Irak, *janjawids* au Soudan, *révolution orange* en Ukraine). Enfin, la quatrième catégorie rassemble des unités qui ont les caractéristiques d'un nom propre car elles désignent un objet (*Naine brune 2M1207* dans le domaine spatial et *Ahoo* dans le domaine automobile), un lieu (*cédraie des Patriarches* au Liban et *Taipei 101* à Taïwan), une opération (*Opération « Grand canal »* et « *Jours de pénitence* »), un organisme (le *Tribunal spécial irakien* ou *TSI* en Irak), une résolution (la *1 559* au Liban), une technique (*Stop and Start* dans le domaine automobile), un engin (*Spaceship One* et *Swift* dans le

⁵ D'après Pruvost et Sablayrolles (2003 : 36), il n'est pas opportun d'exclure des hapax du champ de la néologie évoquant à juste titre le « poids du mot lâché » de Blanche-Noëlle et Roland Grunig.

⁶ Les fiches relatives aux unités choisies comme exemples peuvent être consultées en ligne dans *Néoscope* (www.lb.refer.org/neoscope/projet.php).

⁷ Le mot *chav* a en effet fait son apparition en 2004 dans l'*Oxford Dictionary* où le *chav* est « un(e) jeune personne, souvent sans niveau d'éducation très élevé, qui suit une certaine mode vestimentaire ». Le féminin de *chav* est *chavette*.

domaine spatial), une loi (*Hartz IV* en Allemagne), une manifestation (*Route d'Elissa* dans le domaine sportif), le titre d'un rapport (le *rapport Butler* en Grande-Bretagne et le *rapport Duelfer* aux États-Unis) ou d'un poste (*Monsieur Euro* et *directeur national du Rensei-gnement*), etc.

Ce paramètre lexicographique, malgré son importance, ne suffit pas à lui seul et est complé-té par les trois autres paramètres évoqués par Cabré (1998 : 254) ; il s'agit du paramètre diachronique (apparition récente de l'unité lexicale dans la langue), du paramètre de l'instabilité formelle ou sémantique de l'unité lexicale (ignorance de l'orthographe, de la prononciation ou du sens précis) et, enfin, du paramètre psychologique (une unité lexicale est néologique si les locuteurs d'une langue la perçoivent comme nouvelle). Quoique ces paramètres ne soient pas faciles à cerner pour chaque unité retenue, ils ont servi à justifier certains choix. D'abord, le paramètre diachronique n'est pas simple à circonscrire car la notion de nouveauté est extrêmement éphémère et est perceptible pour les locuteurs pen-dant un temps variable et limité⁸. Étant des néologues prolifiques, les journalistes (tout comme les écrivains et les rédacteurs scientifiques) proposent une création individuelle à un moment donné pour décrire une situation précise ou reprennent à leur propre compte une innovation lexicale créée par une tierce personne. Force est de constater aussi que « les journalistes et des éditeurs de presse choisissent de sélectionner presque systématiquement les passages des discours ou des textes où figurent des néologismes dans les citations qu'ils rapportent ou les extraits radiophoniques ou télédiffusés qu'ils présentent. Le relais qu'ils procurent concourt à une diffusion bien plus grande que celle restreinte à l'auditoire ou au public auquel les néologismes étaient originellement destinés » (Pruvost et Sablayrolles 2003 : 69). Les médias servent ainsi de bassin privilégié aux néologismes, et par le fait de leur omniprésence dans notre monde actuel, ils en assurent une diffusion rapide, voire exer-cent des pressions constantes pour les imposer aux usagers, et ce par une utilisation répétée qui facilite leur éventuelle intégration. Mais, dans un premier temps, plusieurs néologismes peuvent se côtoyer dans un même article pour désigner une réalité nouvellement apparue⁹. De plus, plus la réalité est nouvelle, plus elle est exposée à la synonymie car, dans un but de simplification et de vulgarisation, le terme savant est parfois remplacé par un terme plus usuel sauf s'il est apte à épater la galerie. Et pour produire un meilleur effet, les rédacteurs insistent sur le caractère novateur de la réalité abordée. Dans la base *Néoscope*, ce paramè-tre diachronique fait l'objet d'une note qui mentionne la date d'apparition de l'innovation et son origine.

Par ailleurs, l'instabilité formelle ou sémantique ainsi que le vide lexical ou notionnel ont servi de justification au choix de certaines entrées et ont poussé l'équipe à faire le point sur la situation et à proposer une solution¹⁰. Cette tentative de solution a donné lieu à des néolo-

⁸ D'après Pruvost et Sablayrolles (2003 : 36), « les linguistes ont tendance à restreindre à moins de cinq ans la durée accordée au "sentiment de la néologie" ».

⁹ Les journalistes ont surnommé l'homme de Florès, le *Pilécanthrope* (genre qui regroupait autrefois les *Homo erectus* et *habilis*) ou le *Hobbit* de Florès (en référence au personnage légendaire de l'auteur J.R.R. Tolkien). *Homo floresiensis* est la désignation scientifique.

¹⁰ Le terme *blog*, par exemple, a été retenu même si son existence remonte à plus de dix ans, car, d'une part, la Commission générale de terminologie et de néologie a publié au Journal officiel français du 20 mai 2005 un avis en vertu duquel *bloc-notes* et sa forme abrégée *bloc* remplacent *blog*. D'autre part, un équivalent arabe

gismes « création maison » qui n'ont de valeur que dans la mesure où ils peuvent servir de solution temporaire, soit pour combler le vide existant en attendant une solution définitive, soit pour intervenir avant qu'il ne soit trop tard et que la solution de facilité de l'emprunt ne soit adoptée¹¹.

Quant au quatrième paramètre, le paramètre psychologique, il s'est avéré incontournable, malgré son caractère subjectif, et ce parce que l'extraction ne s'est pas faite automatiquement. En effet, ce qui peut paraître nouveau pour certains, ne l'est pas pour d'autres¹². Ce sentiment de la nouveauté, inhérent au paramètre diachronique, a été aussi exacerbé par les conditions de travail. Ainsi, le fait de coller à l'actualité la plus récente comporte certains risques : d'abord, observer l'actualité à chaud n'a pas permis de prendre le recul nécessaire pour juger des innovations nées d'un besoin urgent de représenter une réalité (choses, idées, faits, etc.) par une désignation, et ce d'autant plus que toute innovation est sujette à des changements tant au niveau de sa désignation qu'au niveau du concept lui-même. En effet, il est rare qu'un néologisme ne fasse pas l'objet de débats avant d'être avalisé. Ainsi, seules l'épreuve du temps et la caution sociale sont susceptibles de chasser les choix erronés (intrus, hapax et mort-nés). Ensuite, il ne faut pas négliger la situation très spécifique du domaine médiatique, à savoir le côté éphémère de certaines innovations qui envahissent les médias et ne tardent pas à disparaître car « la plupart des nouvelles qui occupent le devant de la scène d'actualité sont vouées aux poubelles de l'histoire » (Kientz 1971 : 96). Enfin, le sentiment de la nouveauté comporte le risque de relever des néologismes qui ne le sont pas en réalité dans les domaines de spécialité traités mais qui voient le jour dans les médias au moment où ces derniers décident d'aborder des sujets desquels les profanes sont en général exclus et dévoilent par conséquent les dénominations appropriées. Ainsi, toute nouvelle idée qui commence à circuler dans un premier temps en milieu restreint – le micromilieu des spécialistes – connaît une « seconde naissance » (Pruvost et Sablayrolles 2003 : 38) en entrant dans l'espace médiatique.

À ces quatre paramètres généralement reconnus, nous avons ajouté, en dernier lieu, un cinquième paramètre d'ordre matériel, le paramètre typographique (l'encadrement avec des guillemets et des italiques dans le corps de l'article ou en gras dans les titres) tout en sachant

créé par le procédé d'extension sémantique *مدوّنة*, autre que l'emprunt utilisé *بلوج*, a été répertorié et méritait d'être mentionné. De même, les entrées *spam* et *pourriel* ont été retenues à cause de la profusion des désignations en français et en arabe pour désigner le courrier non sollicité. Dans ce cas, l'équipe a signalé les équivalents existants tout en proposant aussi un équivalent arabe *فمالة* pour remplacer l'emprunt *سبام*.

¹¹ Par exemple, pour l'expression *الفئة الضالة* qui désigne dans la terminologie officielle saoudienne les extrémistes islamistes de la section locale du réseau terroriste d'Oussama Ben Laden, Al-Qaida, dans la péninsule Arabique, l'équivalent français « minorité égarée » figurant en entrée a été proposé par la presse francophone comme traduction littérale du terme arabe tout en notant que l'expression « extrémistes islamistes » y est plus répandue. Cependant, dans un premier temps, l'équipe avait proposé une expression en français qui pouvait mieux refléter le sens spécifique du terme arabe, à savoir « brebis galeuses ». Quant au terme *masstige* – qui désigne un nouveau concept dans le domaine du marketing qui vise à offrir la possibilité à un large public d'accéder à des produits prestigieux, et ce en associant des marques bas de gamme à des étiquettes de prestige, un équivalent arabe *الفخامة للجميع* a été proposé; il traduit le sens de ce terme né de la contraction des expressions *mass market* et *prestige* en anglais.

¹² «Le néologisme dépend d'un jugement relatif et même subjectif, lié à sa définition même, qui repose non pas sur la nouveauté objective, mais sur un sentiment de la nouveauté » (Rey 1988 : 282).

qu'il ne peut être pris en considération de manière systématique¹³. Il indique soit une volonté d'attirer l'attention, soit une « certaine défiance qui se traduit par une mise à distance ». Attirés par les néologismes, les médias ne veulent pas en assumer la responsabilité et se contentent de les transmettre (Pruvost et Sablayrolles 2003 : 70) contribuant ainsi parfois à en assurer la diffusion.

Quoi qu'il en soit et quel(s) que soi(en)t le(s) paramètre(s) pris en compte, le néologisme ne peut exister indépendamment du *corpus* qui justifie sa raison d'être et fournit les traits notionnels qui le caractérisent.

2.2. Analyse du *corpus*

Le *corpus* de référence choisi a comporté, dans un premier temps, un ensemble de quotidiens libanais et arabes choisis en fonction de trois critères : leur type (des quotidiens), leur tirage (les quotidiens à grand tirage) et l'aire géographique qu'ils couvrent (Liban, Syrie, Égypte et Koweït). Cependant, l'équipe a décidé d'en réduire le nombre en ne gardant que les quotidiens à grande diffusion au Liban qui jouissent aussi d'une grande notoriété dans le monde arabe, et ce pour deux raisons : d'une part, certains quotidiens de langue arabe adoptent une politique de rédaction qui se concentre sur une actualité limitée dans un cadre politico-social fermé où l'objectif du journal n'est pas la diffusion de nouvelles mais plutôt un outil de propagande mis au service du régime au pouvoir. D'autre part, la seconde raison est d'ordre pratique : le dépouillement se faisant manuellement¹⁴, le volume de pages à dépouiller par chercheur en l'espace d'une semaine était volumineux.

Une fois la phase de constitution du *corpus* achevée, il fallait se rendre à l'évidence que ce *corpus*, principalement constitué de communiqués et d'articles de presse, ne répond pas à toutes les exigences et ne peut être considéré comme une source fiable de données, et ce pour trois raisons. D'abord, les médias privilégient souvent le côté insolite de la nouvelle en abordant le thème d'actualité dans une approche souvent subjective. En effet, après la collecte des informations, les journalistes veillent à leur mise en forme. Recueillie par un réseau d'informateurs spécialisés (correspondants, reporters, envoyés spéciaux attachés au journal et agences de presse), la matière brute est donc traitée par une équipe de rédacteurs chargée de rendre les messages destinés à la diffusion plus accessibles et attrayants. Ils accordent donc la priorité au sensationnel, filtrent les informations en fonction de leur originalité, facilitent l'intelligibilité en élaguant tout ce qui est trop abstrait et technique, condensent la nouvelle et la cristallisent autour de quelques mots-clés. Par conséquent, dans un texte médiatique, le récepteur attend de l'émetteur qu'il lui apprenne quelque chose au moment où le journaliste suppose toujours que son récepteur est un non-initié. Élaborer une définition à partir de contextes médiatiques ainsi élaborés contribuerait à en faciliter la

¹³ « [...] la corrélation opérée de manière systématique entre l'absence de guillemets et la lexicalisation est contestable, puisque l'on sait qu'un locuteur peut forger un néologisme à son insu en mobilisant (ou non) les ressources du système sans utiliser de marques spécifiques, et qu'inversement une unité lexicalisée qui serait inconnue d'un interlocuteur sera appréhendée par celui-là comme une unité étrangère à sa compétence lexicale et de ce fait reprise éventuellement entre guillemets (Cusin-Berche 2003 : 60) ».

¹⁴ L'extraction automatique n'a pas été envisagée car les quotidiens ne publient en général qu'une sélection d'articles en format électronique. À noter aussi les problèmes liés à l'usage de l'arabe dans les logiciels d'extraction existants.

compréhension à condition que le côté insolite ne prenne pas le dessus sur l'aspect simplificateur.

La deuxième raison qui met doute la fiabilité du *corpus* réside dans les altérations que subit le message et que Keinz (1971 : 77-81) classe en trois catégories : des pertes (*items* manquants), des distorsions (transformation et confusion dans les chiffres et modification des dénominations) et des parasites (le journaliste s'empare d'un élément du message, le développe, brode autour et son volume initial dans le communiqué se multiplie). Donc, la nouvelle reste à la merci de choix qui peuvent avoir des conséquences – parfois néfastes – sur le processus de traitement des données recueillies. Pour faire face à cette situation, le chercheur peut aller au-delà des limites du *corpus* en puisant les données à la source ou dans d'autres sources plus fiables (notamment les sources primaires même si elles sont en anglais). La validité de la définition reste alors tributaire de la qualité et de la quantité des contextes relevés car la définition d'une nouvelle réalité qui n'est pas largement diffusée risque de reproduire les altérations subies, exposant le néologisme à diverses interprétations.

Enfin, la dernière raison qui pousse tout chercheur à prendre ses précautions en traitant un *corpus* médiatique est le facteur temps qui oblige les journalistes à fournir la matière à publier dans les délais les plus brefs, d'où la qualité – parfois douteuse – des propositions terminologiques avancées (notamment quand elles désignent une réalité importée) voire des contextes eux-mêmes. Le seul impératif étant parfois de noircir régulièrement une certaine quantité de papier, la rapidité de production dans la pratique journalistique aboutit parfois à l'émergence à l'improviste de créations (pour combler un vide lexical, proposer une désignation plus correcte, ou se démarquer) dont le sort reste tributaire de leur réemploi dans d'autres contextes.

Simplifications, altérations, rapidité de production sont autant de raisons qui risquent de discréditer le *corpus* choisi. Cependant, il est indéniable que les médias occupent une place prépondérante dans notre univers quotidien car ils servent de relais entre les détenteurs du savoir et du pouvoir et le public en général. Cette fonction d'intermédiaire est aussi assortie d'un devoir, celui d'informer et d'expliquer. D'où l'importance des contextes médiatiques qui encadrent le néologisme pour expliquer son sens, justifier son émergence et en assurer une éventuelle diffusion. Ainsi, tout contexte qui annonce l'émergence d'un néologisme est le support idéal où il est possible de puiser autant de traits notionnels nécessaires à l'élaboration d'une définition. Cependant, une fois que le néologisme a assuré ses différentes fonctions (attirer l'attention du lecteur, le convaincre de l'apparition d'une nouvelle réalité ou combler un vide lexical, etc.), il finit par être utilisé dans les médias comme une unité lexicale ordinaire. Une certaine complicité naît alors entre le quotidien et ses lecteurs car l'équipe de rédaction assume que son lectorat est assidu et est toujours à l'affût de nouveautés. Par conséquent, le contexte qui entoure le néologisme en question se vide progressivement de tout indice qui aide à la construction du sens. Ainsi, la révolution orange en Ukraine, le triangle de la mort en Irak et les états indécis aux États-Unis paraissent désormais dans les différents contextes médiatiques sans que le journaliste ne ressente le besoin de les expliquer. De plus, le *corpus* médiatique n'est pas à l'abri des connotations mélioratives ou péjoratives, notamment quand il s'agit du discours politique. Tenter d'élaborer une définition dans ces conditions est une entreprise ardue surtout si la nouvelle réalité est per-

que selon deux points de vue différents, voire contradictoires, qui se reflètent d'ailleurs dans la désignation elle-même. Par exemple, la grande muraille que le gouvernement israélien est en train de construire à l'intérieur des territoires palestiniens est perçue par les autorités israéliennes comme un moyen d'empêcher des attentats alors qu'elle est perçue du côté des Palestiniens comme une barrière de ségrégation et de séparation, d'où les différentes dénominations adoptées pour la désigner¹⁵.

Témoin vivant des changements de notre monde qui ne peuvent avoir lieu sans que des innovations lexicales ne voient le jour, le *corpus* médiatique est aussi une preuve tangible de la vitalité d'une langue et de la créativité de ses usagers. Constituant la base où il est possible de puiser les emplois spontanés, ce *corpus* sert de filtre qui soumet toute innovation à l'épreuve du temps tout en la rendant accessible au lecteur, contribuant ainsi à en faciliter son éventuelle diffusion.

En conclusion, malgré tous les problèmes rencontrés, il est impossible de nier l'intérêt que revêt l'étude d'un tel *corpus* dans le but de repérer des innovations lexicales qui trouvent une place dans l'actualité et ont des chances ou non d'intégrer le lexique. Face à cette situation, le chercheur, en observateur vigilant, est appelé à être à l'écoute d'un usage en continue mutation et à tenter, dans la mesure du possible, de concilier entre les principes de la terminologie et la réalité sur le terrain. La conception de *Néoscope* est une tentative modeste, dont l'objectif est d'intervenir avant qu'il ne soit trop tard, pour faire état d'un usage déjà en place, ou pour combler les lacunes au niveau de la dénomination ou de la notion, tout en trouvant un compromis avec les exigences des académiciens qui défendent jalousement la qualité de la langue et l'usage *de facto* donc difficilement déracinable. En adoptant la formule de base de données accessible en ligne, l'équipe a misé sur le caractère interactif d'un tel support, dans ce sens où le caractère d'urgence généralement associé à l'émergence des néologismes est pris en compte, tout en offrant la possibilité aux usagers de réagir. Cette structure souple permet aussi de traiter les données *in vivo* en permettant des mises à jour continues. La base de données *Néoscope* peut être considérée en définitive comme un incubateur qui assiste à l'éclosion naturelle des néologismes en observant leurs premiers pas dans l'usage, et en intervenant pour combler toute lacune, et ce, en toute transparence sur la toile, pour en assurer une meilleure diffusion ; la diffusion étant la condition *sine qua non* pour que tout effort terminologique, en général, et néologique, en particulier, ne soit pas vain.

¹⁵ *Mur de l'apartheid, mur de l'annexion et de l'apartheid, clôture de prévention du terrorisme, barrière de sécurité, ligne de sécurité, barrière de séparation, zone de suture* sont les diverses dénominations qui désignent cette muraille. Le cabinet palestinien donne au mur un nom qui fait écho au terme sud-africain synonyme de ségrégation en le nommant le *mur de l'annexion et de l'apartheid* ou tout simplement le *mur de l'apartheid* pour protester contre l'expulsion des Palestiniens et l'annexion de leurs terres. Tandis que pour les Israéliens, il s'agit d'une barrière (ou d'une ligne) de sécurité, appelée officiellement *clôture de prévention du terrorisme* par l'État hébreu car, pour Israël, cette barrière est nécessaire pour se protéger d'incursions terroristes. Cependant, le terme barrière de séparation est employé par ceux qui désirent garder une certaine neutralité par rapport au sujet. Quant au mot *suture* dans la désignation *zone de suture*, il signifie en hébreu tracer un sillon étanche entre les deux peuples.

BIBLIOGRAPHIE

- Awad (G.), 1995 : *Du sensationnel : place de l'événementiel dans le journalisme de masse*, Paris, L'Harmattan.
- Cabré (M.-T.), 1998 : *La Terminologie. Théorie, Méthode et Applications*, Paris-Ottawa, Armand Collin-Les Presses de l'Université d'Ottawa.
- Cusin-Berche (F.), 2003 : *Les Mots et leurs contextes*, Paris, Presses Sorbonne nouvelle.
- Kientz (A.), 1971 : *Pour analyser les médias : l'analyse de contenu*, Paris, Mame.
- Moiran (S.), Bouacha (A.A.), Beacco (J.-C.) et Collinot (A.), dir., 1994 : *Parcours linguistiques de discours spécialisés*, colloque en Sorbonne, 23-25 septembre 1992, Berne, Éditions scientifiques européennes, Peter Lang.
- Pruvost, (J.) et Sablayrolles, (J.-F.), 2003 : *Les néologismes*, Paris, Que sais-je ?, PUF.
- Rey (A.), 1988 : « Dictionnaire et néologie », dans *Terminologie et technologies nouvelles*, actes du colloque tenu à Paris, 9-11 décembre 1985, Québec/Paris, Office de la langue française/Commissariat général de la langue française, pp. 279-289.
- Ross, (L.), 1990 : *L'Écriture de presse : l'art d'informer*, Boucherville, Québec, Gaëtan Morin éditeur.
- CERTA, *Néoscope*, <http://www.lb.refer.org/neoscope/projet.php>

MODERNISER SA LANGUE POUR MIEUX APPRENDRE LES LANGUES ÉTRANGÈRES

Élisabeth Ravaoarimalala

Alliance française

Antananarivo, Madagascar

ravaoarimalala@hotmail.com

***Résumé :** Comment rendre l'apprentissage d'une langue étrangère efficace ? L'apprenant doit être préparé à en accueillir les différents concepts. Disposer d'un outil terminologique bilingue est une priorité. La construction sera basée sur l'exploitation des informations. Pour y faire face, l'usage combiné de l'outil informatique et de la traduction est inévitable. Les produits finaux refléteront et l'interrelation des termes au contexte et la négociation entre experts. Une telle ressource facilitera l'accès aux nouvelles connaissances et par conséquent l'acquisition des langues étrangères.*

***Mots-clés :** didactique de langue, terminologies, traduction, termes et contexte*

1. INTRODUCTION

La position de Madagascar de par le monde l'astreint sinon au plurilinguisme, du moins au bilinguisme, malgache-français, de par l'histoire. Cependant, l'apprentissage du français n'est pas chose facile dans les écoles. Qui plus est, la majorité des élèves n'est bonne ni dans leur propre langue, ni dans la langue étrangère. Comment y remédier ?

L'une des solutions consiste à doter la langue malgache d'outils terminologiques adéquats pour une bonne compréhension des connaissances nouvelles et leur acquisition. Ce qui aura pour effet de la moderniser.

2. LE RÔLE DE LA TERMINOLOGIE DANS L'ACQUISITION DES CONNAISSANCES

Elle conduira l'apprenant des langues étrangères à être conscient des différentes façons d'exprimer les idées d'une langue à une autre. En effet, la manière de concevoir les choses forme un système, qui lui-même est le produit d'un ensemble de facteurs : psychiques, physiques, mentaux, économiques, géographiques, bref, tout ce qui conditionne l'homme dans son environnement. Tous ces facteurs agissent les uns sur les autres et fondent une certaine logique déterminant la vision des choses d'une communauté donnée. Ainsi, si en français, l'orientation du discours porte le plus souvent sur l'agent responsable de l'action, en malgache, elle pourrait aussi bien porter sur l'agent que sur l'objet de l'action, ou la circonstance où celle-ci se déroule. Cette orientation du discours décidera de la forme verbale à utiliser, à savoir, active, passive, circonstancielle. L'usage d'une voix verbale inappropriée conduirait à un non-sens et ne pourrait produire l'effet escompté. La bonne com-

munication dépend du respect de ce cheminement de la pensée qui épouse la logique de la langue elle-même. Ainsi, toute acquisition de nouveaux concepts doit considérer ce processus.

De plus, l'apprenant doit être apte à analyser et à comparer la langue qu'il apprend avec la sienne. Ce qui suppose une bonne connaissance du fonctionnement de sa propre langue. La conscience de la différence sera un pas vers l'acceptation de l'étranger¹ et son intégration en soi.

3. COMMENT AMÉLIORER SES ACQUIS ?

Une fois la structure syntaxique acquise, on avancera dans l'acquisition effective de l'étranger de plusieurs façons :

- par la lecture des romans qui l'habituent aux environnements socioculturels de l'étranger ;
- par la lecture de la presse dont la complexité nécessite l'usage d'un dictionnaire, d'un lexique ou d'une terminologie ;
- par la traduction.

De son côté, l'enseignant doit aider l'apprenant à s'exercer à se faire comprendre par l'emploi des formes admises par tous.

4. L'APPORT DE LA TRADUCTION DANS LE BON APPRENTISSAGE D'UNE LANGUE ÉTRANGÈRE

Nous pensons que la traduction de textes, spécialisés ou normaux, constitue le summum de l'apprentissage de l'étranger. D'ailleurs, cela n'a pu se faire sans l'intervention de la traduction comprise dans le sens d'interprétation et de compréhension.

La traduction des concepts nouveaux passe par une phase explicative où l'incompréhensible devient compréhensible, l'intraduisible devient traduisible, et tient compte du fait que la structure mentale qui façonne les termes² est tributaire des environnements historiques ou socioéconomiques.

Les nouveaux termes apparaissent donc en concomitance avec un événement politique ou social ou autre du pays.

[1] *adin-tsaranga* « lutte de classe », [2] *ny madinika* « les petits » pour indiquer les prolétaires rappelleront la tendance vers le « socialisme » durant la seconde République malgache.

[3] *kolikoly* : « corruption », terme faisant son apparition vers les débuts des années quatre-vingt, lié à la pénurie des produits de première nécessité à Madagascar, phénomène qui a certainement engendré petit à petit la corruption.

¹ Pour dire langue étrangère, nous nous permettons d'emprunter le terme « étranger » de Ricœur

² « terme » est utilisé, ici pour désigner une unité terminologique.

L'évolution des technologies a des incidences indéniables dans les comportements sociaux. Le désir excessif de s'appropriier ces choses pourraient déstabiliser certains, d'où toutes sortes de violences qui ne cessent de se multiplier : vol, viol meurtre...

De cette situation sont nés des termes tels que [4] *manolana* « violer », dont le sens initial est « tordre ».

Toutefois, ces nouveautés ont indéniablement créé plus de bien que de mal.

Exemple, des changements intervenus au niveau de la gestion administrative de l'État font surgir des termes comme : [5] *fanjakana tsara tantana* « bonne gouvernance ».

La lutte intensive contre la propagation du virus du SIDA a engendré l'usage de [6] *fimailo* dont le sens littéral est « moyen permettant de rester vigilant » pour dire « condom ».

Bref, il s'avère alors incontestable que « les termes » soient à traiter dans une dimension multiple intégrant « les aspects cognitifs, linguistiques, sémiotiques et communicatifs » (Cabré 2000 : 14). Ce qui suppose une bonne connaissance de sa propre langue et de l'étranger. Qui plus est, le va-et-vient entre la langue maternelle et l'étranger ne peut être rendu sans l'aide d'une terminologie adéquate. En effet, le traducteur a toujours besoin du travail du lexicologue ou terminologue qui s'est penché sur le sujet. Les « mots » qui s'y trouvent l'aident « à analyser le sens du message et à l'exprimer » (Lederer 1986 : 31).

5. COMMENT ÉLABORER UNE TERMINOLOGIE ?

La préoccupation des lexicographes avait été toujours d'essayer de trouver les correspondances exactes des termes d'une langue à une autre. À Madagascar, le premier recueil de vocabulaires malgaches datait du XVI^e siècle.³ L'effort dans ce sens a suivi un parcours évolutif. Qui plus est, l'entrée massive des connaissances nouvelles et des informations nous pousse à enrichir nos connaissances sur la construction du sens. De plus en plus, et ce grâce à l'apport des nouvelles technologies, il est possible de disposer de textes multiformes pouvant satisfaire un public donné. Le sens des termes utilisés y acquiert un remaniement inattendu de par l'interrelation qu'il entretient avec son milieu environnant. Bien que ce phénomène soit source de richesse, il n'en crée pas moins une certaine divergence appelant à une mise au point des limites de l'usage des termes. Le but étant de développer au mieux la terminologie et la néologie qui favoriseront les activités de traduction vers et à partir d'une langue donnée. Ceci agit dans le sens de la promotion de cette langue et bien plus, de celle du plurilinguisme.

Pour aboutir à cet objectif, nous avons choisi de travailler sur des *corpus* parallèles⁴ que nous avons construits à partir des supports médiatiques, du fait que ceux-ci traitent des sujets différents tels que la politique, l'économie, la culture, l'environnement, l'histoire, la géologie, l'agriculture, la santé, l'éducation. Ces supports ont la probabilité d'atteindre une

³ O. C. Dahl (*cf.* bibliographie).

⁴ « On parle de parallélisme non seulement entre les textes en relation de correspondance traductionnelle, mais aussi en réécriture et réinterprétations de textes monolingues susceptibles de permettre une comparaison » (Zimina-Poirot 2004, 13).

grande partie de la population. Ce qui est un facteur important dans l'évaluation des incidences psychiques des nouveaux termes récoltés. D'après Diki-Kidiri, c'est cette population qui « génère et valide les termes » (Diki-Kidiri 2000 : 31) ainsi que la chance de leur diffusion. Ainsi, avons-nous opté de collecter des articles de journaux traitant du même sujet dans les deux langues, à savoir le malgache et le français, pour y parvenir. Ces articles ont l'avantage d'avoir un certain poids dans l'imaginaire collectif du peuple du fait de leur caractère accessible et méritent pour cette raison d'être pris en compte.

À partir de ces *corpus* parallèles, nous nous sommes proposé d'analyser l'état de la langue malgache pour la rendre capable d'affronter les nouveaux apports de la modernité et profiter au maximum du partage de savoirs dans le contexte de la mondialisation. Malheureusement, nous n'avons pas pu nous étendre dans notre choix. Certains articles sont uniquement en français ou en malgache. Nous nous devons de les traduire pour que l'éventualité de repérage des nouveaux termes ne soit pas négligée.

6. LA PLACE DE LA TRADUCTION DANS L'ÉLABORATION D'UNE TERMINOLOGIE

Outre l'exploitation du *corpus* s'appuyant sur la technique de manipulation d'outils informatiques, nous devons partir de l'idée qu'elle se fonde avant tout sur la traduction.

Celle-ci doit aboutir à la création de termes dont le sens serait « une combinaison de connaissances préalables et de compréhension immédiate de la langue » (Lederer 1986 : 31). À l'exception des termes spécifiquement techniques, nous nous garderons de rendre les mots par les mots, méthode qui aboutira peut-être « à un texte grammaticalement correct mais qui risque de ne pas être intelligible » (Lederer 1986 : 31).

En effet, si nous convenons que la traduction est une réécriture, que le traducteur doit épouser les formes de la pensée de l'auteur, il doit mettre en suvre, forger des constructions pour en donner un effet de sens nouveau qui soit saisissable en même temps par les gens dans le contexte où elles se meuvent. Mieux encore, il a à adopter une attitude flexible lui permettant de s'adapter aux variations de sens que peuvent subir les mots suivant le contexte avec lequel il se trouve confronté. Pour ce faire, il doit avoir tout le temps une attitude réceptive, vis-à-vis de ce que peuvent lui apporter les différents spécialistes pour affiner son analyse.

En somme, « la spécificité » des unités terminologiques « se trouve dans leur aspect pragmatique et dans leur mode de signification. Leur signifié est le résultat d'une négociation entre experts. Cette négociation se produit dans le discours spécialisé à travers des prédications qui déterminent le signifié de chaque unité » (Cabré 2000 : 14)

Cela est rendu possible par le fait que le mot comporte en lui une signification de base qui s'exploite aisément,

– soit de façon totale :

[7] *fahalalahana* « liberté », le sens initial étant « l'ampleur d'un espace ou d'une faculté donnée ».

– soit de façon partielle :

[8] *maridrefy* : indicateur, L'usage initial de *maridrefy* étant « échelle (sur les cartes) » ;
– soit par l'image qu'il insinue et représente :

[9] *sampana* « agence, section, service », dont le sens initial est « branche, embranchement ».

Nous avons pu noter le caractère mobile des unités du lexique qui peut aller du « commun vers le lexique spécialisé » (Cabré 2000 : 14). Tels sont les cas des exemples suivants :

[10] *manotra* « masser » devenu « extorquer de l'argent à quelqu'un » ;

[11] *lomorina* « couvert de mousse ou de moisissure » devenu « prévarié » dans *toetsaina lomorina* « une mentalité pourrie, prévariée » ;

[12] *mason-tsvana* « les yeux (petits trous) des tamis » pour dire « critère ».

Il se produit également une réactualisation de l'usage d'un terme suivi de la spécification de la signification, tel est le cas de *tsikera* « critique, analyse », dont le sens initial était : « action de séparer, de diviser, de détacher, de rendre plus espacé, plus volumineux. »

[13] *tsikera sara* « analyse des coûts »

[14] *tsikera elanelana* « analyse des écarts »

[15] *elanelana* « l'espace, ce qui est entre, intervalle, distance... »

[16] *tsikera andraikitra* « analyse des tâches »

[17] *andraikitra* « charge, devoir, responsabilité »

[18] *tsikera ara-toekarena* « analyse économique »

Ainsi, selon le domaine d'usage du terme, « les conditions discursives » ont activé « seulement quelques informations » le faisant acquérir « une valeur pragmatique particulière, qui est reliée au trait du signifié. L'une de ces valeurs peut être qualifiée de « terminologique » ou de « spécialisée ». De la sorte, le travail consistera à rendre compte de cette valeur et de sa relation avec les autres termes du même système » (Cabré 2000 : 14).

Cela permettra de rendre compte la manière dont est représentée « la connaissance spécialisée » (Cabré 2000 : 14) et par quoi est-elle transmise.

Une telle spécification de la signification peut créer, certes, une homonymie qui n'est que l'expression de la polysémie : « l'homonymie de la terminologie traditionnelle se justifie à notre avis comme la polysémie » (Cabré 2000 : 14).

Pour cette raison, « les unités terminologiques [...] sont à traiter comme des unités discursives provenant d'une seule forme de base : les unités lexicales... » (Cabré 2000 : 14)

Une appropriation de la nouveauté s'est ainsi faite par l'exploitation d'un archétype contenu dans les archives de la mémoire collective. D'après Marcel Diki-Kidiri, « Chaque fois que l'homme est confronté à une réalité nouvelle quelle qu'elle soit, il va spontanément rechercher dans les archives de sa culture les archétypes référentiels qui lui permettront *a priori* d'interpréter cette nouveauté » (Diki-Kidiri 2000 : 28).

Une reconceptualisation de l'idée de « liberté », par exemple, s'est faite dans le choix de *fahalalahana* suivant le mode de pensée malgache qui tend à la comparaison. La liberté est considérée sous l'angle de l'extension de l'espace qui permettra un mouvement sans entrave.

Le terminologue doit donc tenir compte de l'état de la langue, savoir si l'étude pourrait être abordée par le dépouillement des informations sur Internet, ou via des articles de journaux. En effet, comme dit Teresa Cabré : « Un linguiste concerné par la description des unités terminologiques doit les chercher dans les productions orales et écrites des spécialistes ». En somme, le comportement langagier, tel que nous venons de le voir, est indissociable des préoccupations sociales et a des implications sur le sens à donner à des termes tels que : [19] *tranonkala* « site Web », le sens initial étant : « toile d'araignée ».

7. L'ASPECT THÉORIQUE DE L'ANALYSE DES TERMES

Pour parvenir à une étude théorique des termes spécialisés, Teresa Cabré nous propose de partir du texte considéré dans « sa structure et ses conditions discursives », de l'analyser par la suite en vue de détecter les unités porteuses d'une connaissance et d'une signification spécialisées. « Les unités lexicales mises ainsi en évidence sont à analyser encore pour en faire ressortir les relations conceptuelles qui les lient dans le texte, le cas de synonymie – partielle ou totale – afin d'en établir la structure conceptuelle ainsi que sa corrélation avec la structure discursive » (Cabré 2000 : 15).

Cet ensemble constitue l'analyse des unités terminologiques dans une théorie à base linguistique, mais ne peut rendre compte de leur spécificité significative en tant que « véhicules de connaissances spécialisées » (Cabré 2000 : 15).

Bref, ajoutons que la considération du terme avec le texte où il se trouve ainsi que le contexte où il s'insère va permettre :

- d'éviter une simplification des concepts qui risque d'en donner une vision caricaturale, et faire perdre ainsi le contenu du sujet ;
- d'avoir des visions parcellaires de la chose afin de trouver un terme qui traduit le concept d'une façon compréhensible sans pour autant sacrifier l'information de départ. Faire de cela une règle absolue. En effet, « On doit pouvoir évaluer avec les locuteurs natifs aussi bien qu'avec les experts du domaine, la marge de tolérance des écarts éventuels entre le microsystème source et le microsystème cible (auquel on aboutit), l'acceptabilité, la flexibilité et la facilité d'emploi des termes retenus pour composer le microsystème cible » (Diki-Kidiri 2000 : 31) ;
- de toujours se remettre en question, bien réfléchir sur ce qu'il y a derrière les mots, même les plus courants, dans leur relation avec les autres termes de la phrase et l'idée du contexte ;
- de réviser toutes les connaissances en contexte politique, géographique et autre, aussi bien au niveau local qu'international ;

- Diki-Kidiri préconise de collecter tous les concepts susceptibles de rendre le concept étranger que l'on veut s'approprier. Et il ajoute qu'une « analyse critique des mots candidats récoltés est nécessaire, aussi bien du point de vue de leur formation, de leurs connotations, que de leur adéquation technique aux concepts de la langue source que l'on veut rendre dans la langue cible ; là où plusieurs synonymes ou variantes étaient proposés ou possibles. Ceci constitue la dernière opération de la production terminologique elle-même » (Diki-Kidiri 2000 : 31).

L'objectif fondamental consistera ensuite à valider les termes pour aboutir à l'instrumentalisation de la langue cible.

En fait, le but est de parvenir à la constance du concept que l'on a de la chose. Il s'agit de rechercher « une expression rendant clairement la pensée qui fera apparaître celle-ci dans toute sa précision à ceux qui ne peuvent l'appréhender qu'au travers de la traduction » (Lederer 1986 : 33).

Cette démarche exclut tout ce qui crée la confusion, telle que la redondance, les ambiguïtés, la synonymie du point de vue de l'expression, bref, le manque d'une stricte précision. Ensuite, « les unités de signification spécialisée acquièrent leur valeur terminologique dans un domaine de spécialité quand elles sont effectivement utilisées dans ce domaine » (Cabré 2000 : 15).

C'est une visée qui, nous l'avons déjà souligné plus haut, requiert la participation des instances interdisciplinaires.

Marcel Diki-Kidiri pense que « La méthodologie de la recherche exige un travail pluridisciplinaire dans plusieurs aires culturelles. La recherche associera donc le concours de linguistes mais aussi de spécialistes des domaines traités (médecins, informaticiens, entomologistes, géographes, juristes, historiens, etc.). Enfin, la recherche tiendra compte éminemment des recherches fondamentales actuelles dans le domaine de la cognition et de l'anthropologie culturelle pour mieux affiner sa propre démarche en terminologie » (Diki-Kidiri 2000 : 31). Cela requiert continuellement, de notre part, le développement d'un esprit de recherche valorisant qui soit axé sur l'appliqué.

Bien entendu, une langue est toujours extrêmement riche. Aussi, n'est-il pas toujours possible d'employer tel mot à la place de tel autre. Cela ne fait que confirmer cette nécessité d'approfondir le concept que comporte le terme utilisé, en interrelation avec les éléments qui l'avoisinent. Ces nuances doivent être senties par l'établissement des terminologies. Le but étant de faire que la langue puisse faire passer des choses tant en termes d'émotions que d'informations. Ce qui s'obtiendra par une approche exhaustive des textes et contextes auxquels est rattaché le terme, objet d'étude. Ce qui nous fera aboutir à une banque d'informations qui fera sans cesse appel à d'autres.

8. CONCLUSION

Les langues des puissances mondiales prédominent. Les langues des pays minoritaires comme Madagascar ne perdent pas pour autant de leurs valeurs. Qui plus est, le malgache est toujours une langue de communication quotidienne comprise par tous, c'est la langue de

dix-sept millions d'habitants⁵, environ, dont les non instruits constituent les 45,7 %⁶. Le réel développement du peuple repose sur une très bonne acquisition des informations scientifiques et technologiques. Ce qui ne saurait se réaliser sans le passage de cette connaissance par la langue du pays. D'où la nécessité impérieuse d'arriver à traduire les apports de la modernité en langue du pays pour que les gens ne passent pas à côté de ces connaissances. En outre, nous pensons que la tendance au bilinguisme malgache-français est un fait inéluctable. D'ailleurs, la langue de l'administration et des médias est majoritairement le français. Il se présente un déséquilibre de communication, entraînant celui du partage des savoirs et des informations. Dès lors, comment rétablir cet équilibre ?

Disposer d'une langue à la hauteur des nouvelles technologies s'avère primordial. Une traduction qui apporte le plus de lumière possible y contribuera. L'objectif étant de parvenir à une langue qui soit « capable de couvrir plus de domaines d'expériences et de connaissances humaines » (Diki-Kidiri 2000 : 29). Les incidences indéniables seront la meilleure connaissance des concepts véhiculés par les informations, et un apprentissage des langues étrangères plus rapide et performant.

BIBLIOGRAPHIE

- Cabré (M. T.), 2000 : « Terminologie et linguistique : la théorie des portes » dans *Rifal* n° 21, pp. 10-15.
- Dahl (O. C.), 1951 : *Malgache et Maanjan, une comparaison linguistique*, Oslo, Egede-Instituttet.
- Diki-Kidiri (M.), 2000 : « Une approche culturelle de la terminologie » dans *Rifal* n° 21, pp. 27-31.
- INSTAT (Institut national de la statistique), 2003, *Enquête auprès des ménages 2002, rapport principal*. Direction des statistiques des ménages, Antananarivo.
- Lakoff (G.) et Johnson (M.), 1985 : *Les Métaphores dans la vie quotidienne*, traduit de l'américain par Michel Deformel avec la collaboration de Jean-Jacques Lecerle, Paris, Les Éditions de Minuit.
- Lederer (M.), 1986 : *Interpréter pour traduire*, Paris, Didier Érudition.
- Mounin (G.), 1963 : *Les Problèmes théoriques de la traduction*, Éditions Gallimard.
- Rabenoro (I.), 2004 : « Bien apprendre le malgache pour bien apprendre les langues étrangères » dans *Annales de la faculté des lettres et sciences humaines*, Université d'Antananarivo, n° 13.
- Rajaonarimanana (N.), 2000, « la lexicographie malgache ; état des lieux et perspectives » dans *Dictionnaires bilingues méthodes et contenus*, Paris, Honoré Champion.

⁵ Selon la projection démographique, variante moyenne : (d'après l'information qui nous a été donnée) le dernier recensement indique 12 210 000 habitants et date de l'année 1993, le prochain aura lieu en 2007.

⁶ Selon l'INSTAT : les non instruits sont ceux qui n'ont jamais fréquenté l'école ou qui ont atteint au maximum la troisième année du primaire.

À titre d'information, ajoutons que le niveau primaire, c'est-à-dire ceux qui ont achevé au moins la 4^e année du primaire (classe de 8^e), et n'ont pas dépassé la classe de 4^e des collèges, est de 42,4 %.

Le niveau secondaire, formé de ceux qui ont achevé au moins la classe de 3^e des collèges et au plus une classe terminale, est de 9,3 %.

Le niveau supérieur, formé de ceux qui ont obtenu le baccalauréat ou ont déjà fréquenté un établissement d'enseignement supérieur est de 2,5 %.

- Ravaoarimalala (E.), 2004 : *La Néologie lexicale dans le malgache contemporain. L'Évolution de la langue malgache vue à travers l'étude des néologismes*, thèse dirigée par Narivelo Rajaonarimanana, pour l'obtention du doctorat nouveau régime en études africaines (4 tomes), Paris, INALCO.
- Razafimanjato (J.), 1993, *Recensement général de la population et de l'habitat, projections et perspectives démographiques à Madagascar*, vol. II, Rapport d'analyse, t. VIII, INSTAT (Institut national de la statistique), Antananarivo.
- Ricœur (P.), 2004 : *Sur la traduction*, 2^e tirage, Paris, Bayard.
- RR. PP. Abinal et Malzac, s.j., 1888 : *Dictionnaire malgache-français*, Paris, Éditions Maritimes et d'Outre-mer.
- Seleskovitch (D.) et Lederer (M.), 1986 : *Interpréter pour traduire*, Paris, Didier Érudition.
- Zimina-Poirot (M.), 2004 : *Approches quantitatives de l'extraction de ressources traductionnelles à partir de corpus parallèles*, thèse dirigée par André Salem, Université Paris III – Sorbonne nouvelle – école doctorale, langage et langues.

LA RECONNAISSANCE AUTOMATIQUE DES NÉOLOGISMES DE SENS

Salah Mejri

laboratoire de linguistique informatique (LLI)

CNRS, Université Paris XIII

Villetaneuse, France

salah.mejri@lli.univ-paris13.fr

Résumé : *La détection automatique des néologismes de forme ne pose actuellement que peu de problèmes. Pour qu'il en soit de même pour les néologismes de sens, il faut une approche appropriée qui fournit à la machine des données formelles dégroupant tous les emplois d'une unité lexicale. En nous appuyant sur les travaux menés dans le cadre du LLI, nous fournissons un modèle de détection automatique des néologismes de sens.*

Mots-clés : *néologisme, sens, traitement automatique*

1. NÉOLOGISMES, FORMES NON RECONNUES ET DÉTECTION AUTOMATIQUE :

Pour parler de reconnaissance automatique, il faut disposer d'une définition clairement établie permettant de distinguer d'une manière univoque les objets à détecter, en l'occurrence les néologismes. Si, par ailleurs, nous intégrons dans le même objet des distinctions internes, il ne faut en aucune façon que ces distinctions portent préjudice à la définition de l'objet. Le résultat, c'est qu'il faut disposer d'une définition englobant toutes les configurations du même objet. Dans le cas de la néologie, il faut que la définition englobe à la fois les néologismes de forme et les néologismes de sens.

Nous proposons la définition suivante :

x est un néologisme dans Z , si $x \notin Z$ à t (Z étant l'ensemble des unités lexicales [formes et emplois] de la langue étudiée ; t : le repère temporel choisi pour l'étude)

Une telle définition, pour qu'elle soit opératoire, exige au moins trois conditions :

- l'inventaire systématique de toutes les formes appartenant à Z à l'instant t ;
- un moteur de reconnaissance des formes ;
- des filtres effectuant le tri de toutes les formes non inventoriées, qu'elles relèvent des variantes normées ou des formes déviantes, parmi lesquelles il y a lieu de distinguer les fautes, les formes ludiques et les néologismes.

Les inventaires, extensibles et actualisables, sont couramment employés dans la correction orthographique : l'ordinateur ne reconnaît pas les formes qui n'existent pas dans le dictionnaire. Cette reconnaissance n'est possible que lorsqu'on dispose d'un moteur de reconnais-

sance, lequel peut être enrichi par des fonctions par lesquelles on peut obtenir des formes jugées correctes. Cette dernière fonction repose sur des calculs de probabilité fondés sur les points de ressemblance entre la forme déviante et les formes qui partagent avec elle un certain nombre de points communs. Comme ce calcul n'est pas actuellement très fin, on a souvent affaire à plusieurs propositions.

On peut dire que les services rendus par ces moteurs sont relativement satisfaisants et donnent des résultats relativement fiables pour les néologismes de forme. S'il y a encore un vrai problème, il se situe au niveau des filtres qui doivent faire la distinction entre les formes non reconnues.

Devant des exemples comme *événement*, *évènement* et *évenement*, le programme sera capable d'indiquer que les deux premières formes sont correctes et que la dernière ne l'est pas. Il peut même fournir des indications supplémentaires sur la deuxième forme, si cette fonction y est intégrée. Mais il ne serait pas capable de fournir la moindre indication sur la dernière tant qu'au moins un filtre approprié n'est pas fonctionnel. Tout ce qu'il peut faire, c'est de dire qu'elle n'est pas inventoriée.

Il est clair que de ce point de vue notre définition peut donner satisfaction. Comment donc faire pour l'appliquer à tous les objets de la néologie ? Pour apporter des éléments de réponse à cette question, il faut évaluer les descriptions linguistiques disponibles dans la matière.

2. L'APPROCHE STRATIFIÉE

Les descriptions courantes s'inscrivent dans une approche qu'on peut considérer comme stratifiée. Elles sont fondées sur un découpage faisant de chaque niveau de l'analyse linguistique une discipline autonome : phonologie, morphologie, syntaxe, sémantique, lexicale, etc. Appliquée à la néologie, cette démarche rencontre beaucoup de difficultés dans la reconnaissance automatique.

Le plan morphologique est le seul à être opérationnel même s'il est isolé du reste des niveaux. Mais il faut tout de suite préciser que cela n'est vrai que de la morphologie strictement lexicale. Même si beaucoup d'efforts ont été fournis pour apporter des éléments de réponse pour tout ce qui relève de la syntaxe, comme les accords, il y a encore du chemin à faire. Les descriptions actuelles sont limitées à l'unité lexicale, indépendamment de ses contextes. La machine ne reconnaît que la forme telle qu'elle figure dans la base. Tout ce qui concerne la relation entre les unités relève d'un autre niveau, celui de la syntaxe.

Au niveau syntaxique, il n'y a pas de définition précise d'un néologisme syntaxique. Si on en a une, elle ne peut se concevoir dans l'optique qui est la nôtre que d'une manière strictement formelle pour que la machine soit capable de reconnaître les formes en question. Même avec une telle définition, pour qu'un programme soit capable de reconnaître automatiquement les néologismes qui ont comme contrepartie une trace syntaxique comme le changement de construction (un verbe transitif devient intransitif ou l'inverse, un transitif direct se construit indirectement ou l'inverse, un transitif indirect change de préposition, les cas de conversion qui touchent toutes les unités passant d'une catégorie grammaticale à une autre, etc.), il faut disposer d'une description exhaustive de tous les emplois de tels items

lexicaux. Cela ne peut pas se faire indépendamment des autres niveaux. Les descriptions dont nous disposons dans les dictionnaires nous fournissent des éléments se présentant comme suit :

[1] Exemples de conversion :

aérien :

$\Sigma 1$: adj. De l'air, de l'atmosphère

$\Sigma 2$: n.m. Antenne

commercial :

$\Sigma 1$: adj. Qui a rapport au commerce

$\Sigma 2$: n. Personne chargée des relations commerciales dans une entreprise

Sur le plan sémantique, toutes les descriptions disponibles prennent appui sur l'analyse sémique. Un mot change de signification si sa structure sémémique change, et ce moyennant tous les mécanismes sémantiques décrits dans les traités : métaphore, métonymie, synecdoque, extension de sens, restriction de sens, jeu de synonymie et d'antonymie, etc.

[2] Exemples de changement de signification :

– *vert* : l'Europe verte, le franc vert, la révolution verte

$\Sigma 1$: qui n'est pas mûr, se dit des végétaux qui ont encore la sève

$\Sigma 2$: de la nature, de la campagne (*une restriction accentuelle*)

– suicidaire :

$\Sigma 1$: qui mène, qui tend au suicide

$\Sigma 2$: qui mène à l'échec, à la faillite (*passage du concret à l'abstrait*)

– invivable :

$\Sigma 1$: très difficile à vivre

$\Sigma 2$: où il est difficile, voire impossible de vivre (*on passe de l'objet au lieu*)

Sur le plan strictement lexical, les travaux disponibles couvrent des aspects relatifs aux domaines d'emplois comme les champs lexicaux développés par Louis Guilbert, aux niveaux de langue, aux particularismes régionaux, aux archaïsmes, à l'étymologie, etc. Comme on le constate, ces différents aspects peuvent avoir une efficacité certaine dans la reconnaissance automatique puisqu'ils correspondent à des catégories formelles, mais ils ne le sont réellement que quand ils sont corrélés à d'autres paramètres.

Si nous consultons le *Grand Robert*, nous remarquerons que pour des unités comme *base*, *souris* et *virus*, il est fourni des indications qui aident avec une relative précision à la détermination de la signification de l'item en fonction du domaine : pour le mot *base*, il est indiqué qu'il appartient au langage militaire (*les bases militaires, couper une armée de ses bases, base de ravitaillement*, etc.), à la géométrie (*la base d'une pyramide, d'un cube, d'un*

triangle, etc.), à la chimie (*l'eau est une base*), aux mathématiques (*la base du système décimal est dix*), à la linguistique (*base d'un mot*), à l'informatique (*base de données*), etc. On peut faire le même constat pour *souris* et *virus* : dès qu'on en spécifie le domaine, on discrimine les emplois d'une manière satisfaisante. Le virus en biologie n'est pas le même en informatique ; la souris mammifère n'est pas la souris de l'ordinateur.

L'inconvénient majeur de ce genre de description, c'est la stratification. Le fait de limiter la description à un seul niveau d'analyse réduit considérablement les capacités de détection des formes recherchées. Il est possible pour un humain d'établir les connexions nécessaires entre toutes les informations disponibles à chaque niveau et de repérer l'emploi recherché. Une machine en est incapable. Si on reprend chaque niveau, on peut apprécier les limites imposées dans chaque cas.

Sur le plan morphologique, les limites de la performance des descriptions de forme deviennent saillantes dès qu'il y a identité de forme entre deux mots ou deux emplois, comme c'est le cas pour *livre* (poids et document), *abattage* (d'arbres, de minéraux, d'animaux, etc.), etc. Il en est de même dans les distinctions sémantiques : comment formaliser les variations sémiques pour que la machine soit capable de les identifier ?

Nous croyons que la solution réside dans une approche qui abandonne le postulat selon lequel l'unité minimale de signification est le mot. Même si le mot peut être un point de départ à l'analyse, sa signification ne peut pas être conçue indépendamment du contexte dans lequel il est employé. Ce contexte peut avoir plusieurs configurations. Il est d'abord phrastique, et c'est le niveau de base qui détermine la valeur réelle de toute unité lexicale. S'y ajoutent les domaines, les niveaux de langue et l'environnement contextuel.

Concrètement, cela signifie qu'il serait beaucoup plus efficace, dans une perspective de traitement automatique, d'abandonner l'approche stratifiée pour une approche intégrée, où l'on fournit dans la même description toutes les données qui soient capables de conduire à des solutions univoques. Prenons l'exemple du verbe *prendre*. *Le Grand Robert* y consacre 43 définitions différentes. Si on tient compte des nuances ajoutées à l'intérieur de chaque signification, on peut facilement doubler ce chiffre. Si nous prenons les exemples suivants :

[3] Luc a pris un cachet d'aspirine, *absorber*

[4] Luc a pris une cigarette, *fumer*

[5] Luc a pris son vélo, *utiliser*

[6] Luc a pris un thé, *boire*

[7] Luc a pris un sandwich, *manger*

[8] Marie a pris le malfaiteur, *attraper*

[9] Luc a pris l'autoroute, *emprunter*

[10] Luc a pris l'avion, *voyager*,

on remarquera qu'il y a autant de significations que d'emplois. C'est ainsi qu'on peut apprécier l'importance d'une démarche qui intègre l'item lexical dans son cadre phrastique.

L'approche stratifiée ne peut être appréciée que dans la perspective de l'apprentissage humain des langues. L'intelligence humaine a cette caractéristique : elle est capable d'établir de multiples connexions entre des entités appartenant à des strates différentes et hiérarchisées. Tel n'est pas le cas de la machine pour le moment. C'est ce qui justifie épistémologiquement le recours, dans la perspective du traitement automatique des langues, à l'approche intégrée qui tente de fournir à la machine toutes les indications relatives aux emplois propres à chaque unité lexicale.

3. LA DESCRIPTION INTÉGRÉE

Nous nous limitons à quelques indications :

- l'unité minimale d'analyse est la phrase ;
- la phrase élémentaire est faite d'un prédicat ;
- chaque prédicat détermine un schéma d'arguments ;
- chaque emploi représente une entrée ;
- la description doit comporter toutes les indications utiles pour dissocier les emplois : l'actualisation, les restructurations, le domaine, le niveau de langue, etc.

La réalisation d'un dictionnaire électronique capable de discriminer les emplois ne peut se faire que sur la base de la description des emplois de chaque item lexical, qu'il soit prédicatif ou pas.

Si on appliquait cela au verbe *prendre*, on obtiendrait des descriptions comme :

- prendre/N0 : hum/N1 < logement > S : occuper, loger en
- prendre/N0 : hum/N1 < voie > S : emprunter
- prendre/N0 : hum/N1 < aliment > S : manger
- prendre/N0 : hum/N1 < médicaments > S : absorber
- prendre/N0 : hum/N1 < boissons > S : boire
- prendre/N0 : hum/N1 < argent > S : prélever

Cette description part de l'item lexical *prendre* et en décrit l'emploi en termes de classes d'objets, c'est-à-dire la classe des items lexicaux figurant dans les positions syntaxiques du schéma d'arguments et devant être lexicalement saturées. Il s'agit dans l'exemple choisi de descriptions en intension. Pour qu'elle soit opérationnelle, elle doit être complétée par une description en extension. Cela se fait par l'énumération exhaustive de tous les items de toutes les classes du schéma d'arguments de chaque entrée.

Pour effectuer un tel travail, il faut détailler les grandes classes comme *humain* qui ne sont en réalité que des simples traits sémantico-syntaxiques s'opposant à d'autres comme *concret*, *animé*, etc. Le fait d'affiner l'analyse de la grande classe des humains, conformément à leurs emplois, nous permettrait d'y voir un grand nombre de classes plus précises

partageant des emplois ayant un fonctionnement identique. Nous en fournissons un exemple tiré des travaux du LLI :

- *adepte* : protestant, taoïste
- *âge* : enfant, vieillard
- *ape* (appellatif) : monsieur, sire
- *col* : foule, troupe
- *défmor* (défaut moral) : menteur, voleur
- *défphy* (défaut physique) : difforme, boiteux
- *défpsy* (défaut psychologique) : imbécile, stupide
- *eth* (ethnie) : Sioux, Peuls
- *êtremyth* : sylphe, farfadet
- *fon* (fonction) : adjoint, arbitre
- *grade* : capitaine, colonel
- *hum* (humain proprement dit) : homme, gens
- *ins* (instrumentiste) : violoniste, pianiste
- *loc* (locatif) : Parisien, Biterrois, Allemand
- *locm* (locatif par métonymie) : rue, salle
- *lscol* (lieu scolaire) : école, faculté
- *lspect* (lieu de spectacle) : cinéma, théâtre
- *ltr* (lieu de travail) : usine, bureau
- *malphy* (maladie physique) : tuberculeux, cancéreux
- *malpsy* (maladie psychique) : névrosé, fou

Il en est de même d'autres classes :

- logement : appartement, studio, chambre, villa, pied-à-terre, bungalow, etc.
- voie : route, autoroute, piste, sentier, pente de montagne, montée, etc.
- aliment : chocolat, gâteau, fromage, pizza, biscuit, etc.
- etc.

Il faut préciser que cette démarche inscrit les descriptions linguistiques dans une perspective expérimentale. Puisqu'elle est fondée sur des données empiriques fournies par le *corpus*, elle est conçue dans une logique d'ouverture, qui permet de réactualiser régulièrement toutes les données.

Si nous combinons les descriptions en *intension* et les descriptions en extension, théoriquement, on obtient toutes les phrases possibles. Le simple croisement d'une ligne des descriptions en intension fournies plus haut avec les classes décrites en extension donnerait lieu à un nombre impressionnant de phrases possibles, et ce indépendamment des variations en rapport avec les phénomènes d'actualisation et de restructuration :

- *prendre/N0 : hum/N1 < voie > S : emprunter*
- les humains en position de N0 :
 - *adepte* : protestant, taoïste...
 - *âge* : enfant, vieillard...
 - *col* : foule, troupe...
 - *eth* (ethnie) : Sioux, Peuls...
 - *hum* (humain proprement dit) : homme, gens...
 - *loc* (locatif) : Parisien, Biterrois, Allemand...
 - *voie* : *route, autoroute, piste, sentier, pente de montagne, montée, etc.*

Les phrases possibles seraient le fruit du croisement des listes établies des classes dans le cadre du schéma d'argument. Nous aurions par exemple :

prendre/N0 : hum (protestant, taoïste, enfant, vieillard, foule, troupe, Sioux, Peuls, homme, gens, Parisien, Biterrois, Allemand...) /N1 < *voie* (route, autoroute, piste, sentier, pente de montagne, montée...) >S : *emprunter*. Il suffit d'y ajouter les actualisateurs (déterminants et expression des catégories du temps, du mode, de l'aspect, de la personne...) pour obtenir des énoncés normalement constitués.

4. MÉTHODE

Au lieu d'avoir un dictionnaire dans sa version traditionnelle, on aurait un dictionnaire dont les entrées seraient des phrases comportant toutes les descriptions en intension. Un adjectif comme *ingrat* aurait une description comme celle-ci :

adj	catadj	N0	N1	syn	ant
ingrat	sol : fertilité	loc (sol)		stérile	fécond, fertile
ingrat	appr.vue : laid	(visage, physique)		déplaisant, disgracieux	avenant, plaisant
ingrat	comport.envers autrui	hum	envers, pour, vis-à-vis de hum	oublieux	reconnaissant

adj	catadj	N0	N1	syn	ant
ingrat	appr.trav.	ina (métier, travail)		difficile, pénible	

où l'on indique la catégorie sémantique à laquelle appartiennent le premier argument (N0), le deuxième argument (N1), la nature des prépositions si la construction de l'adjectif l'exige, le synonyme dans chaque emploi ou l'antonyme s'ils existent. S'il s'agit d'un verbe, on fournirait d'autres indications comme c'est le cas avec *adhérer* :

adhérer/N0 : inc/N1 : à inc/N2 : /Ad : /Nm : adhérence/Sy : coller/D : tec/

adhérer/N0 : hum/N1 : à ina < opinion >/N2 : de hum/Ad : /Nm : adhésion/Sy : être d'accord avec/D : psy/

adhérer/N0 : hum/N1 : à hum < collectif >/N2 : /Ad : adhérent/Nm : adhésion/Sy : être adhérent/D : pol/

où l'on indique toute la structure argumentale avec les classes auxquelles appartiennent les paradigmes figurant dans chaque position. Dans le cas de ce verbe, le premier emploi s'oppose aux deux derniers par une structure argumentale qui implique des inanimés concrets et qui relève du domaine de la technique. Les deux derniers se distinguent entre eux sur la base des domaines – l'un relève du psychologique, l'autre du politique – mais également grâce à la classe de N1. Cette différence entre le premier emploi et les deux derniers est corroborée par une différence de nominalisation : *adhérence* pour le premier emploi et *adhésion* dans les deux autres. On peut avoir les mêmes descriptions pour les noms, les prépositions et les adverbes.¹

Tout ce que nous avons dit concerne la prédication du premier ordre. La prédication du second ordre implique une hiérarchie entre les prédicats, qu'ils soient actualisés ou pas (G. Gross et M. Prandi 2004).

Avec ce type de dictionnaire, on n'est plus dans la perspective de l'analyse stratifiée où le sens est décrit en termes de sèmes difficiles à détecter, mais où les entrées phrastiques dégroupent sur la base de la méthode décrite tous les emplois du même item : autant d'emplois, autant d'entrées. Concrètement, cette démarche donne aux changements de sens une dimension formelle. Si dans un *corpus* quelconque se trouve une phrase qui ne s'insère pas dans les entrées décrites en intension et en extension dans le dictionnaire, trois possibilités s'offrent à nous :

- soit la phrase en question relève d'un emploi effectif de l'un des *items* impliqués par la description de leurs emplois dans le dictionnaire électronique, il s'agit alors d'une incomplétude ou d'une erreur de description. Dans ce cas, nous procédons aux rectifications nécessaires. Cet aspect illustre parfaitement la démarche expérimentale qui

¹ Au LLI, on procède à la description systématique des prédicats et des classes d'objets.

sort la linguistique des généralisations hâtives et inscrit la recherche dans les données empiriques où le *corpus*, nécessairement ouvert, corrige les descriptions effectuées et les valide en même temps ;

- soit la phrase est d'un emploi franchement erroné, c'est-à-dire inacceptable et ne répondant pas aux critères retenus pour apprécier la bonne formation des énoncés phrastiques : dans ce cas, notre dictionnaire peut proposer des corrections conformes aux descriptions disponibles dans sa mémoire. On passe ainsi de la correction formelle cantonnée dans les limites de l'item lexical, couramment utilisée dans la fonction orthographique, à la correction de nature combinatoire impliquant à la fois les dimensions lexicales, syntaxiques et sémantiques ;
- soit la phrase est tout simplement déviante. La différence avec la catégorie précédente, c'est que ce genre d'emplois relève d'une acceptabilité contextuelle. Dans ce cas, on cherche à quel niveau le nouvel emploi se situe. Nous prenons le cas le plus simple, celui de la variation argumentale. Un prédicat dont le schéma d'arguments n'admet qu'un paradigme de classes précises peut à n'importe quel moment de son emploi se voir attribuer de nouvelles classes. Il s'agit là de néologismes de sens. Le dictionnaire est censé les détecter. Un travail ultérieur consistera à les intégrer dans une description plus fine de la langue.

5. LA CONFIGURATION GÉNÉRALE DU SYSTÈME DE RECONNAISSANCE AUTOMATIQUE :

Deux méthodes sont possibles :

- l'approche modulaire : elle consiste à choisir une unité lexicale et à en faire une description exhaustive capable de rendre compte de toutes les phrases possibles avec l'item en question. Cela signifie qu'on combine la description en intension avec la description en extension ;
- l'approche générale consiste à décrire des paradigmes généraux traduisant une structuration quelconque du lexique, comme celle des prédicats d'affect², tout en optant pour le croisement des deux types de description.

Techniquement, cette méthode est actuellement réalisable dans le cadre modulaire. Des travaux sur le verbe *prendre* sont actuellement en cours. Son objectif est de fournir toutes les phrases élémentaires possibles avec cet item lexical : emplois supports, emplois prédicatifs, emplois figés, emplois tropiques, non tropiques, etc.

Des tests ont été réalisés sur des unités beaucoup moins polysémiques comme *brouillé*.

² Nous procédons actuellement dans le cadre des équipes du LLI à une description plus économique des prédicats, qui consiste à décrire les racines prédicatives au lieu d'en faire des descriptions séparées. Si des prédicats comme *respecter*, *respect* et *respectueux* se partagent la même racine prédicative, il est plus économique de décrire directement la racine et lui rattacher les fonctionnements nominaux, verbaux et adjectivaux de ce prédicat.

1. *brouillé* météo : ciel et temps/évé (temps, ciel)/ chargé, couvert, menaçant/clair, radieux
2. *brouillé* clair-confus/ina (idées, esprit)/ confus, troublé/clair, lucide
3. *brouillé* compétent : fâché avec/hum/avec ina/fâché
4. *brouillé* état phys.pass. : dérangé/hum (estomac)/ barbouillé
5. *brouillé* rapports entre hum : fâché avec/hum/avec hum/en froid, fâché
6. *brouillé* teint : pâleur/inc/plombé, terreux/éclatant, frais

L'application de cette description à l'un des *corpus* sur lesquels nous travaillons (Frantext, journaux, Toile, etc.) donne toutes les occurrences qui vérifient cette description effectuée et fournit des emplois qui sortent de ce canevas (*cf.* un échantillon choisi d'une manière arbitraire [la quatrième page de notre relevé] en annexe sur les emplois de *brouillé*). Comme on le constate, pour cet adjectif, la liste du premier argument est la suivante : *image, piste, scène politique, époque, fin de siècle, carte politique, ligne, mémoire, évocation, politique*. Une fois confronté aux descriptions en intension, il s'avère clairement que des emplois entrent difficilement dans les classes répertoriées. Ne figurent pas dans cet échantillon les emplois 1, 3, 4, 5 et 6. Il ne reste que le cas 2 qui exige comme premier argument un inanimé abstrait. La liste que nous avons, moyennant quelques précisions, répond à ce critère. L'intérêt de cette extraction réside dans l'établissement de listes exhaustives des classes. Plus les listes sont complètes, plus la détection automatique est précise. Dans le même relevé, nous avons repéré des emplois figurant en annexe 2 où le premier argument entre difficilement ou est plus ou moins éloigné de la typologie établie.

Le fait qu'on y croise des classes relativement hétérogènes allant d'emplois peu distants à des emplois qui rompent franchement avec les classes établies rend nécessaire la présence d'un filtre. Il aura pour fonction :

- d'améliorer les descriptions des emplois disponibles,
- de faire figurer dans les entrées les emplois non listés,
- de retenir les nouveaux emplois et de les inscrire dans des fichiers propres pour rendre le système plus prédictible.

Partant des exemples retenus, il est clair que l'emploi le plus marqué, c'est l'emploi nominal dans l'exemple suivant :

[11] Les irrégularités, les incidents de parcours, le *brouillé*, le brouillon se substituent aux plénitudes antérieures de l'encre.

Arrivent des emplois qui impliquent des items appartenant à des classes relativement éloignées des classes fréquentes : *tennis, quête sonore, clivage*, etc.

Pour résumer, nous donnons la configuration générale du système :

- [1]. Description en intension en termes de classes d'objets des items lexicaux ;

- [2]. Description en extension de toutes les classes impliquées par l'emploi des items étudiés ;
- [3]. Validation par le *corpus* ;
- [4]. Confrontation des descriptions avec les données du *corpus* ;
- [5]. Le résultat : reconnaissance des emplois non listés dans le dictionnaire électronique ;
- [6]. Filtrage des occurrences pour isoler les néologismes de sens et les emplois qui enrichissent la description effectuée.

Actuellement, un travail en cours cherche à systématiser cette technique qui, si elle est menée à bien, améliorerait énormément la détection automatique des néologismes de sens.

BIBLIOGRAPHIE

- Gross (G.), 1984 : « Étude syntaxique de deux emplois du mot coup », dans *Linguisticae Investigationes* VIII, 1, pp. 37-61.
- Gross (G.), 1995 : « Une sémantique nouvelle pour la traduction automatique : les classes d'objets », dans *Tribune des industries de la langue et de l'information électronique*, 17-18-19, pp. 16-19.
- Gross (G.), 1999 : « Élaboration d'un dictionnaire électronique », dans *Bulletin de la Société linguistique de Paris*, XCIV, 1, pp. 113-138.
- Gross (G.) et Prandi (M.), 2004 : *La Finalité. Fondements conceptuels et genèse linguistique*, Duculot, Bruxelles.
- Gross (G.), 2005 : « Un dictionnaire électronique des adjectifs du français », dans *Cahiers de lexicologie*, 86-1, pp. 11-33.
- Mejri (S.), 1990 : « Néologie et variétés lexicales », *Visages du français. Variétés lexicales de l'espace francophone*, coll « Actualité scientifique », John Libbey, Eurotext/AUPELF-UREF, Paris-Londres 1990, pp. 11-26.
- Mejri (S.), 1995 : *La Néologie lexicale*, Publications de la faculté des lettres de la Manouba, 1995, série linguistique, volume IX, 380 pp.
- Mejri (S.) 2000 : « Figement lexical et renouvellement du lexique : quand le processus détermine la dynamique du système », dans *Le français moderne*, LXVIII, 1, CILF, 2000, pp. 39-62.
- Mejri (S.), à paraître : « Figement, néologie et renouvellement du lexique », numéro spécial de *Linx*, sous la direction de Salah Mejri, Gérard Petit et Geneviève Petiot, Hommage à M.-F. Mortureux.
- Pruvost (J.), 2000 : *Dictionnaires et Nouvelles Technologies*, PUF.
- Pruvost (J.), 2002 : *Les Dictionnaires de langue française*, PUF, « Que sais-je ? », n° 3622.
- Pruvost (J.) et Sablayrolles (J.-F.), 2003 : *Les néologismes*, PUF, « Que sais-je ? », n° 3674.
- Sablayrolles (J.-F.), 2002 : « Fondements théoriques des difficultés pratiques du traitement des néologismes », dans *Revue française de linguistique appliquée*, VII-1.
- Sablayrolles (J.-F.), 2003 : *L'Innovation lexicale*, Champion.

ANNEXE 1

Il arrive à l'heure du sprint final avec une image *brouillée*, trouble, incertaine.

La piste fut longtemps *brouillée*, car ce n'était pas une mais trois études d'opinion qui allaient être publiées le même jour.

Mais l'image de la CFDT était sortie de cet épisode à nouveau *brouillée*.

Ainsi la scène politique corse est-elle plus *brouillée* que jamais, hésitant entre le vieux et le neuf, oscillant entre alliances et mésalliances.

Dans un monde qui va comme il va, dans une époque aussi *brouillée*, aussi changeante et finalement aussi opaque que la nôtre, nous pouvons être tentés de porter sur scène des attitudes de révolte, de sarcasmes ou au contraire de retrait hautain.

J'espère que seront rejetés avec elle le goût irresponsable des polémiques indues et l'image affligeante et *brouillée* que ne méritent ni l'État, ni la démocratie, ni les élus que vous êtes.

Toutes ces questions, il est naturel que les intellectuels – donc les revues – se les posent en cette fin de siècle *brouillée*, où tous les repères s'estompent.

L'image de la médecine et des médecins est désormais *brouillée*.

Il y révisé une carte politique *brouillée*, contemple les méfaits de la télé, observe une France moyenne qui a voté Le Pen parce que la droite et la gauche les ont laissés s'enfoncer.

Soucieuse de lutter contre les pénuries de main-d'œuvre qualifiée qui frappe ces filières, l'UIMM entend s'appuyer directement sur des jeunes en sortie de collège et essayer de les attirer vers ces métiers dont l'image est toujours *brouillée* et dévalorisée, explique M. Dominique de Calan, directeur de la formation de l'UIMM.

Cette ligne singulière est aujourd'hui un peu *brouillée*.

Vous ajoutez la corruption et la misère, cela aboutit à une mémoire *brouillée*, un syncrétisme crétin, des mots vidés, stupides ;

Denis Clarac et Philippe Hauer, étudiants à l'École d'art d'Aix-en-Provence, proposent, eux, un montage vidéo qui compose par la saturation des signes (photos, graphiques, images, sons, voix...) une évocation *brouillée* et désenchantée de la monstruosité du siècle.

Plusieurs responsables locaux déplorent l'image *brouillée* que renvoie désormais le MDC.

Or, la politique de Lionel Jospin dans le conflit kosovar est très *brouillée*.

ANNEXE 2

Souvent invité, il a pu ainsi faire décoller son nouvel album, *Le Pont Mirabeau*, qui allait cahin-caha et reconstruire une identité un peu *brouillée*.

La frontière, cependant, entre les mesures de police classiques et l'arsenal antiterroriste s'est révélée dans le débat passablement *brouillée*.

Il n'a pas manqué d'amateurs de principes généraux, Élie Faure ou André Malraux, pour découvrir en lui le champion du style linéaire, style du Sud minéral et lumineux opposé à la peinture nordique, forcément *brouillée* et ténébreuse.

M. Mitterrand leur a ouvert un boulevard, mais il a tout de même du temps devant lui pour rendre à la politique française à l'Est une cohérence aujourd'hui *brouillée*.

Les proies sont généralement sous-évaluées, dirigées par des équipes vieillies qui n'en tirent pas un bénéfice maximal et à la stratégie *brouillée*.

Comment imaginer que le Christ, nouvel Adam, aurait pu parler une langue *brouillée* plutôt que l'idiome adamique des premiers temps ?

Sur scène, le groupe, augmenté de deux musiciens, met les guitares en avant, emporté par Stuart Braitwaite, s'élance dans une longue quête sonore, *brouillée* et bruitiste.

Restituer cette littérature, délicieusement *brouillée* comme l'horizon d'un matin de nordet, demande autant de patience et de poésie que pour attraper en plein vol nuptial un Orsenna mâle tournoyant dans le faisceau du phare de Rosédo.

Les irrégularités, les incidents de parcours, le *brouillé*, le brouillon se substituent aux plénitudes antérieures de l'encre.

Mais le clivage entre les deux camps, – conservateur et libéral – est bien plus *brouillé* que par le passé.

Mais pour que le jeu du bipartisme (républicain démocrate) ait quelques chances d'être un tant soit peu *brouillé*, il faut qu'un candidat indépendant ait beaucoup d'argent et soit populaire, au moins dans une grande région.

Elle s'est surtout donné les moyens de battre les deux joueuses les plus accrocheuses parmi les dix meilleures, Patty Schnyder et Dominique Van Roost, grâce à un tennis plus intransigeant moins *brouillé* de fautes...

Cette hésitation s'explique en grande partie par une perte des repères politiques, pour une génération qui a grandi dans un paysage idéologiquement *brouillé*.

CRITIQUEZ VOS DICTIONNAIRES BILINGUES... À BON ESCIENT

André Dussart

Institut supérieur de traducteurs et interprètes (ISTI)
Haute École de Bruxelles (HEB)
1180 Bruxelles, Belgique

Résumé : La communication analyse les critiques traditionnelles adressées aux dictionnaires bilingues et propose une grille d'évaluation sur la base de critères autres que l'absence de l'un ou l'autre terme ou la difficulté de choisir parmi la liste des équivalents présentés. Les critères retenus sont la pertinence de la sélection, la systématisme du vocabulaire, l'efficacité et l'évidence du contexte, la précision des explications, la plausibilité des traductions et l'assistance fournie pour l'usage des mots et des syntagmes.

Mots-clés : exhaustivité, pertinence, systématisme, univocité, efficacité

1.

Les traducteurs connaissent bien les reproches classiques adressés aux dictionnaires bilingues : le mot recherché ne s'y trouve pas, car il est trop nouveau ou trop spécifique ; il est mal traduit, ce qui arrive ; une grande liste d'équivalents est présentée sans explication, commentaire ou définition, de sorte que le choix de la bonne traduction relève du jeu de loterie ; enfin le classement alphabétique masque l'éventuelle structuration ou arborescence des concepts, (Bretthauer 2 000 : Heft 3, 121). La discussion qui va suivre découle d'une expérience déjà longue de plus d'une décennie avec le *Dictionnaire de la technique industrielle, allemand-français/français-allemand* du Dr Ernst. La présente communication se propose de passer en revue ces critiques fréquentes résultant d'expériences malheureuses avec des dictionnaires bilingues.

2.

L'absence d'un mot n'est nullement rédhibitoire dans un dictionnaire polytechnique : Michael Grade (2002 : Heft 2, 49) estimait le nombre de termes techniques à 50 millions en 1995, parmi lesquels 8 millions de termes de chimie, 3 millions de termes médicaux et 3 millions de termes d'électrotechnique-électronique. Même dans un glossaire hyperspécialisé, le lexicographe est-il jamais sûr de disposer d'un *corpus* d'extraction exhaustif ? Divers produits sont fabriqués par plusieurs firmes, chacune ayant ses procédés et parfois son vocabulaire (pour l'*ABS*, le système d'antiblocage des roues, il existe des différences entre les techniques des différents constructeurs automobiles). L'arborescence fournit un moyen théorique de vérifier l'exhaustivité du vocabulaire, mais en pratique, pour les éoliennes par exemple, il faut s'assurer que tous les éléments constitutifs sont envisagés : le mât, les pales

de rotor et leur calage d'angle, le contrôle et la régulation de la vitesse, la turbine etc. La synonymie vient compliquer encore la tâche : *wind turbine*, *wind turbine generator*, *wind generator*, *wind power plant*, *wind power station*, *wind energy conversion system*, *wind-wheel*, *wind mill*, *wind machine* sont synonymes ou quasi-synonymes (Wagner 2000 : Heft 3, 108).

L'exhaustivité semble donc illusoire, car même absolue, elle est éphémère : de nouvelles normes sont édictées, des découvertes sont faites, l'évolution des produits et des techniques a des implications sur le vocabulaire. Cette quête de l'exhaustivité absolue renvoie aux *corpus* de dépouillement totalement ou partiellement informatisés récoltés avec circonspection sur l'internet (dans les sites des fabricants) et dans les supports traditionnels (livres et revues).

En effet, les *corpus*, aussi vastes soient-ils, ne sont que des échantillons. Ils doivent être variés, car il est risqué de se limiter à une seule revue technique dans un domaine. Il vaut mieux se méfier des tournures stylistiques particulières, des effets de mode, des idiotismes de certains rédacteurs. À ces exigences vient s'ajouter celle de la représentativité : les *corpus* doivent refléter correctement la langue de spécialité, par conséquent la compétence de l'auteur et l'intention du texte (vulgarisation, manuel d'entretien, texte s'adressant uniquement à des spécialistes) entrent également en ligne de compte.

Un critique reproche l'absence de la « chaîne du froid » dans le dictionnaire Ernst, tout en se réjouissant de la présence de l'« avion furtif ». Il serait facile d'inverser la démarche et d'objecter que l'avion furtif est du domaine militaire et apparaît surtout dans des revues très spécialisées ou épisodiquement dans la presse en cas d'incident ou de conflit. On pourrait, par conséquent, également estimer que l'avion furtif n'est pas vraiment à sa place dans un dictionnaire de la technique industrielle.

D'après Hoffmann (1987 : 147), on peut estimer que le nombre de termes vraiment fréquents d'un domaine technique ou scientifique varie de 1 000 à environ 5 000 (médecine, chimie, physique), selon le pourcentage de fréquences relatives retenu. L'utilisateur du dictionnaire s'intéresse assez peu aux fréquences : il rencontre aussi des mots ou syntagmes plus rares et il s'indignera facilement de leur absence. On peut évidemment lui suggérer de consulter des dictionnaires très spécialisés, tout en sachant que l'offre en la matière est très riche pour l'anglais et moins abondante pour d'autres langues.

Un autre écueil de la statistique est la création quasi permanente de mots nouveaux. S'il est aisé de calculer le nombre total de mots et de locutions rencontrés dans l'œuvre de Goethe ou de Victor Hugo et donc d'établir des fréquences relatives, il est beaucoup plus malaisé d'y arriver dans un ensemble ouvert à l'innovation, tandis que l'obsolescence menace d'autres mots, soit parce que la technique disparaît (les cartes perforées en informatique, les disquettes ou *floppy disks*), soit parce qu'un terme en remplace un autre (normalisation, proposition de traduction d'un anglicisme, effet de mode etc.). Comme le volume acceptable d'un dictionnaire incite les éditeurs à limiter le nombre de pages, des termes disparaissent d'une édition à l'autre, au grand dam des historiens des sciences ou de scientifiques confrontés à des évocations de techniques actuellement abandonnées mais importantes dans l'évolution d'une fabrication.

Pour concilier les exigences d'une démarche scientifique et les contraintes de temps disponible, il faut sélectionner des sous-domaines très délimités. Il serait aberrant de vérifier si le dictionnaire traite correctement les énergies renouvelables dans leur ensemble. Le domaine (cellules photovoltaïques, fours solaires, éoliennes, usines marémotrices, énergie de la houle et des vagues...) est encore trop vaste. On pourrait éventuellement se limiter aux éoliennes, sélectionner les termes fréquents ou indispensables en définissant strictement le contenu de ces notions de statistique. Il suffit alors de chercher les entrées correspondantes dans le dictionnaire. Bien évidemment, il n'est pas question de se limiter à un seul sous-ensemble dans l'évaluation d'un dictionnaire polytechnique. L'idéal est de contrôler le traitement de domaines variés et d'alterner des technologies de pointe (l'optoélectronique, la microélectronique, les télécommunications) et des métiers plus traditionnels (la verrerie, la métallurgie, l'industrie du cuir). Au total cinq ou six domaines devraient permettre un jugement fondé.

En résumé, pour analyser la qualité d'un dictionnaire, il convient de sélectionner quelques domaines très restreints, d'établir une liste de termes fréquents ou jugés importants et d'en vérifier la présence ainsi que la qualité de leur traduction. Certains termes ou phrasèmes n'apparaissent que rarement et seulement dans des contextes bien particuliers (citons la machine à facetter les alliances en bijouterie).

Il conviendrait sans doute de privilégier la systématique du lexique plutôt que l'inaccessible exhaustivité. La terminologie classique postule l'organisation des connaissances dans une structure conceptuelle souvent représentée sous forme d'arborescence. Dès lors, les termes reflètent cette structuration et s'organisent facilement. C'est oublier que la langue, même les langues de spécialité, s'inscrivent dans la vie sociale, donc dans l'évolution et que les accidents de l'histoire la marquent profondément. Cette méthode connaît donc également des limites : elle ignore les contextes (l'usage, la phraséologie, la synonymie) et elle privilégie le référent.

3. QUEL TYPE DE CONTEXTE EXPLICATIF POUR QUEL DICTIONNAIRE ?

L'interprétation de la signification des mots s'opère par la discrimination (identification des référents d'un terme/lexie, distinction entre les homonymes) et par la délimitation (filtrage de certains aspects du sens). Les unités de signification n'ont pas toutes une identité stable, mais elles sont souvent des agglomérats imprécis. De plus, elles subissent des variations de sens et d'usage au cours du temps.

L'explication doit être ciblée sur un public. Le public visé par un dictionnaire de technique industrielle est très vaste et comprend à la fois des experts, des étudiants ou spécialistes en devenir, des traducteurs et même des lecteurs de journaux (la presse relate que des pirates de l'air ont coupé le transpondeur de l'avion, parle d'un avion furtif, de lancement de fusées ou navettes spatiales, de systèmes *GPS*, d'explosifs etc.).

Il est par conséquent difficile de cerner les attentes d'un public aussi diversifié : le lecteur de journaux souhaite seulement connaître l'équivalent d'un terme étranger dans sa langue maternelle. D'autres se pencheront sur la nature de l'équivalence (totale ou partielle),

d'autres encore souhaiteront trouver le contexte morphosyntaxique d'usage du terme et les domaines d'application, pour la rédaction dans une langue étrangère.

Face aux contraintes d'espace disponible dans le livre – tandis que les supports informatiques offrent des capacités de stockage énormes – et aux contraintes de temps alloué au travail, le lexicographe mettra certains moyens en œuvre. La définition ou définition partielle est gourmande en espace mémoire ou papier, elle s'impose rarement. De plus, la définition complète n'a généralement qu'un intérêt anecdotique pour le profane. Le contexte définitoire s'extrait rarement tel quel du *corpus*, le plus souvent, c'est un métatexte bref, même un énoncé incomplet, créé à partir du *corpus* ou d'autres sources de référence :

[1] *Bubblejetdrucker, m. (thermischer Tintenstrahldrucker, bei dem eine kleine Flüssigkeitsmenge erhitzt wird, wobei eine eingeschlossene Luftblase sich vergrößert und die Tinte nach außen schleudert)*, imprimante à jet d'encre thermique à bulles d'air.

Contrairement à l'encyclopédie, le dictionnaire bilingue se contentera de quelques aspects discriminatoires de la définition : si pour le *Petit Robert* (2003), la vandoise est « un poisson d'eau douce (cypriniformes), très commun » et pour le *Grand Dictionnaire encyclopédique Larousse* (1986) « un poisson téléostéen des eaux douces, d'Europe occidentale à la Sibérie, voisin du chevaine, mais plus élancé et au ventre argenté... », une définition discriminatoire suggérera, par exemple, « poisson de rivière de couleur argentée, brillante, au corps très allongé. Se distingue du petit chevesne par sa nageoire anale concave et du gardon par l'iris de l'œil clair, ainsi que par la ligne latérale très incurvée » (Balzat-Dussart, 1978, 44). Cette dernière définition est loin d'être complète et se limite à certains aspects bien particuliers, elle est suffisante pour l'identification, alors que les deux précédentes sont trop vagues. La difficulté est donc de sélectionner un minimum de critères (sèmes) dans l'ensemble, pour garantir une identification correcte du référent.

Le recours à l'antonyme ou terme opposé est très rare : il convient de noter qu'il ne s'agit pas souvent d'une véritable opposition au sens logique du terme :

[2] *Kopfbremssystem (Gegensatz : Heckbremse), n.*, système de frein à l'avant (moulinet de pêche)

[3] *Grundlastkraftwerk (Gegensatz : Spitzenkraftwerk), n.*, usine électrique de base (pour la charge de base), centrale de base ; f.

[4] *Teilblattverstellung (Gegensatz : Ganzblattverst.), f.*, variation du pas d'une partie de la pale

La référence à un holonyme (pour la définition, voir M.-C. L'Homme 2004 : 98) est, quant à elle, relativement fréquente ; elle marque une relation d'appartenance ou d'inclusion :

[5] *Saugkrümmer, m.*, coude d'aspiration (pompe)

[6] *Hochleistungsschnecke, f.*, (Plast), vis à grand rendement (extrudeuse)

[7] *Verschlussschraube, f.*, (Geschütz), vis de culasse

Les synonymes ou paronymes tellement décriés par certains linguistes, parce que la synonymie « n'assure pas la décomposition sémantique du défini » (Melčuk *et al.* 1995 : 82) sont bien utiles dans la plupart des cas, même si la véritable synonymie est rare :

[8] *Förderpumpe*, f., quasi-synonyme *Baggerpumpe*, f., pompe de dragage ; f.

Parfois une note technique spécifiant un aspect du référent (opération, produit [9], forme de la pièce [10], mesures de poids, fréquence, altitude [11, 12, 13], débit, vitesse ou fonction de l'objet [14, 15, 16, 17]) précise l'usage du terme :

[9] *Schalenguss*, m., coulée en coquille (opération) ; fonte en coquille (produit)

[10] *Hakenschraube (j-förmig gebogen)*, f., vis à crochet

[11] *Seam (Partie von 54,431 kg)*, n., bloc de 54,431 kg (verre)

[12] *Ultraschall (20 bis 100 kHz)*, m., ultrason

[13] *Stratopause (50-55 km Höhe)*, f., stratopause

[14] *Abdrückschraube für Räder*, f., vis de chasse

[15] *Schnellbauschraube für Gipskarton*, f., vis à fixation rapide

[16] *Schneidstahl für Ausdreharbeiten*, m., burin d'alésage

[17] *Spülschlamm-Pumpe (Rotary- Bohren)*, f., pompe à boue

La référence à une norme est moins explicite et nécessite des recherches supplémentaires en cas de doute :

[18] *Ringschraube (DIN 580)*, f., vis à anneau de levage

[19] *Monierzange (DIN 5242)*, *Rabitzzange* ; f., *Flechtermasse* ; f., tenailles russes

Le commentaire est parfois utile, pour souligner le statut du terme : les contextes régionaux, les hapax legomena, les figures de style, les anglicismes, les termes vieilliss, rares, recommandés ou à proscrire :

[20] *Madenschraube (Südd.)*, f., vis sans tête

[21] *Pendler*, m., navetteur (Belg.), banlieusard (abus) ; m.

[22] *Pin-Grid-Gehäuse*, n., boîtier fakir (terme le plus courant), boîtier matriciel (terme officiel) ; m., matrice de broches ; f.

Les indications morphologiques (déclinaisons, conjugaisons) ne trouvent pas leur place dans le dictionnaire technique, même si certains le déplorent. En revanche, certaines notes syntaxiques se révèlent précieuses, par exemple pour mentionner l'emploi d'une préposition :

[23] *getaktet [mit 300 MHz]*, cadencé [à 300 MHz]

[24] *herunterladen [auf einen Rechner]*, télécharger des fichiers [dans un ordinateur]

Certains lexicologues sont intransigeants : ils estiment que le dictionnaire bilingue devrait permettre la rédaction dans les deux langues et, par conséquent, fournir des syntagmes, des

idiomes, des collocations, des stéréotypes phraséologiques (Mel'čuk *et al.* 1995 : 45). Il serait facile d'objecter que même les meilleurs dictionnaires unilingues ne permettent pas à un étranger de rédiger parfaitement : imagine-t-on un Allemand ou un Anglais écrivant le français avec la seule aide du *Petit Robert* ? Oui, s'il maîtrise déjà très bien le français.

Toutefois, nombreux sont les syntagmes nominaux traduits, ce qui n'étonnera guère, puisque la déverbalisation de la langue technique est un phénomène souvent mis en évidence : le verbe est supplanté dans son rôle prédicatif par des tournures nominales (groupes de substantifs accompagnés d'adjectifs).

Enfin, le dictionnaire devrait attirer l'attention du lecteur sur les termes protégés. Certaines firmes veillent jalousement à la mention des brevets, que les auteurs de dictionnaires sont priés d'omettre, parce qu'il n'est pas possible de retrouver la trace de tous les termes et sigles brevetés de par le monde et que les éditeurs désirent ainsi se prémunir contre d'éventuels procès. Si une firme pouvait prouver que certains brevets sont signalés, tandis que les siens manquent, elle pourrait exiger des dommages et faire retirer le livre du commerce. Le seul clin d'œil au lecteur est l'usage d'une majuscule : Plexiglas, caoutchouc Mousse, fermeture Éclair, Crochet X, éclairage Scialytique (en salle d'opération).

4. LES CRITÈRES D'ÉVALUATION

Cette brève analyse nous mène à la définition de critères d'évaluation et d'amélioration du dictionnaire : le critère de pertinence, le critère de systémativité, le critère d'efficacité, le critère de précision, le critère de vérité/plausibilité, le critère d'usage.

4.1. Le critère de pertinence : l'inclusion d'un élément dans la liste doit se justifier, mais il est souvent difficile de distinguer les termes pertinents et non pertinents, d'autant plus qu'un terme très général peut prendre un sens particulier dans un contexte donné. Même des mots utilisés assez souvent dans la presse ou les journaux télévisés sont introuvables dans bien des dictionnaires généraux. Ainsi le verbe hélitreuiller (*heraufwünschen*), fréquent dans tous les contextes de sauvetage, n'était répertorié dans aucun dictionnaire français-allemand généraliste en 1999, bien que l'opération d'hélitreuillage soit de plus en plus fréquente en cas d'accidents d'alpinisme, de naufrage, d'inondations ou d'incendies. Il est donc entré dans le dictionnaire Ernst de la technique industrielle, bien qu'il ne relève pas à proprement parler de la technique industrielle. La pertinence est une exigence très subjective ; de plus, elle est fonction des objectifs du dictionnaire et du public visé.

4.2. Le critère de systémativité est rencontré, si le lexicographe a réalisé une collecte et un dépouillement de documents sérieux et de qualité, dont il a tiré un inventaire précis des mots selon une échelle d'importance relative au domaine traité. Si la systémativité a pour objet de limiter les effets du hasard dans les choix opérés et d'éviter les lacunes graves, elle n'est pas fondamentalement le reflet des fréquences statistiques ; en effet, un terme pivot dans la structuration d'un domaine pourrait apparaître relativement rarement dans certains *corpus*, qui ne sont jamais que des échantillons plus ou moins représentatifs.

4.3. Le critère d'efficacité du contexte explicatif est respecté, si l'information fournie au lecteur est simple et évidente par elle-même et nécessite un effort minimal de compréhension. Les illustrations ou photos sont bien entendu un support utile, mais parfois d'un maniement délicat : elles sont gourmandes en espace et souvent protégées par des droits d'auteur. Ce critère suppose l'adéquation au public visé et la sélection très stricte des explications. De ce point de vue, l'information peut rester partielle et incomplète : le dictionnaire bilingue n'est pas une encyclopédie du savoir.

4.4. Le critère de précision correspond à l'exigence de discrimination entre les entrées du dictionnaire. L'utilisateur sera à même d'attribuer une traduction correcte au mot inconnu, après avoir levé toutes les ambiguïtés éventuelles d'une longue liste. Ainsi, selon les contextes, *Steg* sera un étai dans la construction, un cloisonnement de colombage, l'aile médiane d'un profilé en T, l'entretoise reliant des pièces de machine, un peigne de filière, le sous-pied d'un pantalon de ski, un barreau de chaise, une barre de cintre, un embarcadère, une passerelle pour piétons, une barrette d'accumulateurs, un porte-satellite de boîte de vitesses... Heureusement, la plupart du temps, les unités ont une identité plutôt stable dans les textes techniques et leur signification varie moins en fonction des conditions de la communication que dans la langue de tous les jours.

4.5. Le critère de vérité ou plutôt de plausibilité implique la vérifiabilité et par corollaire la fiabilité du dictionnaire : l'explication sert d'ailleurs aussi à vérifier la plausibilité.

4.6. Le critère d'usage vient compléter les précédents : l'auteur doit aider l'utilisateur en lui fournissant des indications syntaxiques (l'usage des prépositions, par exemple) et en lui signalant le statut du terme proposé : recommandé, conseillé, néologique, vieilli, régional, abusif, proscrit. Quant aux idiomes, aux collocations, ils font généralement défaut dans un dictionnaire polytechnique bilingue, mais les possibilités actuelles et futures de numérisation, de mémorisation et d'analyse automatique de textes laissent espérer des progrès importants dans un avenir proche.

5. En conclusion : L'idéal que se fixent les lexicologues est louable, mais actuellement inaccessible, en raison des contraintes matérielles liées à la préexistence de la banque de données lexicales datant de plusieurs décennies (faut-il l'abandonner et tout recommencer à zéro ?), en raison des contraintes de productivité dans un délai assez court (imposé à un seul auteur, les crédits manquant pour rétribuer des équipes), en raison des contraintes d'espace disponible, en raison des contraintes d'accès à des sources diversifiées et sûres ainsi que des contraintes liées aux carences du logiciel de saisie. Par ailleurs, il faut laisser au lecteur, au rédacteur et au traducteur une part de travail, car même dans un système automatisé, c'est toujours l'opérateur humain qui garde le contrôle. Personne n'attend qu'un dictionnaire soit riche et parfait au point d'expliquer à lui tout seul un poème de Mallarmé ou de Rimbaud. Où serait le plaisir de la lecture ?

BIBLIOGRAPHIE

- Baldinger (K.), 1984 : *Vers une sémantique moderne*, Paris, Klincksieck.
- Balzat (N.), Dussart (A.), 1978 : *Guide de la pêche en Ardenne*, Gembloux, Duculot.
- Béjoint (H.), 2002 [1994] : *Modern Lexicography, An Introduction*, O.U.P.
- Bergenholtz (H.) et Tarp (S.), 1995 : *Manual of Specialised Lexicography, The Preparation of Specialised Dictionaries*, Amsterdam-Philadelphia, John Benjamins.
- Brethauer (P.), 2000 : *Das zweisprachige Fachwörterbuch : die Krücke des Übersetzers*, Lebende Sprachen, Berlin : Langenscheidt Verlag, Jahrgang 2000, Heft 3, pp. 121-124.
- Ernst (R.), 1999 : *Wörterbuch der industriellen Technik, Band III, Deutsch-Französisch, 5. vollkommen überarbeitete und erweiterte Auflage von Prof. Dr André Dussart*, Wiesbaden, Brandstetter Verlag.
- Ernst (R.), 2003 : *Wörterbuch der industriellen Technik, Band IV, Französisch-Deutsch, 5. vollkommen überarbeitete und erweiterte Auflage von Prof. Dr André Dussart*, Wiesbaden, Brandstetter Verlag.
- Gouadec (D.), 1990 : *Terminologie, Constitution des données*, Paris, AFNOR gestion.
- Grade (M.), 2002 : *Auswirkungen des wachsenden naturwissenschaftlich-technischen Wissens auf Beruf und Ausbildung technischer Fachübersetzer*, Lebende Sprachen, Berlin, Langenscheidt Verlag, Jahrgang 2002, Heft 2, pp. 49-56.
- Hoffmann (L.), 1987 : *Kommunikationsmittel Fachsprache, eine Einführung*, Berlin, Akademie-Verlag.
- L'Homme (M.-C.), 2004 : *La Terminologie, principes et techniques*, Montréal, Les Presses de l'Université de Montréal.
- Melčuk (I.), Clas (A.) et Polguère (A.), 1995 : *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve/Duculot/AUPELF-UREF
- Polguère (A.), 2003 : *Lexicologie et sémantique lexicale, notions fondamentales*, Montréal, Presses de l'Université de Montréal.
- Pulitano (D.), 2003 : *Ein Evaluationsraster für elektronische Wörterbücher*, Lebende Sprachen, Berlin, Langenscheidt Verlag, Jahrgang, 2003, Heft 2, pp. 49 et sqq.
- Rondeau (G.), 1984 : *Introduction à la terminologie*, Boucherville, Gaëtan Morin.
- Wagner (I.), 2000 : *Windkraft-Glossar (DE-EN)*, Lebende Sprachen, Berlin, Langenscheidt Verlag, Jahrgang 2000, Heft 3, pp. 107-116.

DU CONTEXTE À LA CITATION : LES RÉCENTS DÉVELOPPEMENTS DE LA DICTIONNAIRIQUE

Jean-Nicolas Desurmont

Bruxelles

jdesurmont@yahoo.fr

Résumé : *À partir d'un corpus de dictionnaires récents de langue française (essentiellement du XX^e siècle) nous signalerons certains phénomènes de la pratique citationnelle. Notre réflexion propose à la fois une synthèse de la réflexion actuelle sur la citation dans la littérature métalexigraphique et des propositions théoriques novatrices. Nous avons cru bon, dans le cadre de ce colloque, d'ouvrir notre propos sur des aspects plus généraux de l'étude de la citation de telle manière à aborder les aspects méthodologiques et historiques bouleversés par l'usage de l'informatique.*

Mots-clés : *lexicographie française (histoire), métalexigraphie, régionalisme, linguistique (histoire), citation*

1. BILAN DES ÉTUDES

La métalexigraphie récente s'est davantage intéressée à la définition qu'à la citation. Un bref bilan des études sur la citation ne saurait omettre l'étude de Josette Rey-Debove (1971) sur la sémiotique des dictionnaires qui consacre une partie importante à l'exemple dictionnaire¹ en y dressant une typologie, ni celle de Sidney Landau qui offre une vision plus proche de la rédaction lexicographique. Ce ne sont là que deux exemples d'études sur la citation et l'exemple, dont les approches divergent. Mais la tradition métalexigraphique française regorge de monographies qui comportent une section ou un chapitre consacré à la citation. Rey-Debove, dont la terminologie va marquer les études métalexigraphiques, utilise sa terminologie métalinguistique telle *chose-nommée* (signifié des noms), *signe-nommant* (unité du lexique) et *mot-entrée* (adresse) afin d'y analyser le contenu et la sémiotique de la pratique citationnelle². Alain Rey offre quant à lui une courte synthèse sur les exemples dans les dictionnaires des années 1967 à 1977 (Rey 1977 : 134-152)³ en s'intéressant à l'équilibre entre *exemples signés* et *exemples forgés* (nous reviendrons sur ces termes) de même qu'au fait qu'il s'agisse ou non de syntagmes, d'exemples forgés non pertinents sur le plan stylistique, etc. L'analyse des citations dans les dictionnaires de langue française des années antérieures a été effectuée en partie par Bernard Quemada (1968), Josette Rey-Debove (1971), Georges Matoré (1968) et plus récemment dans le n° 12-13 de la revue *Lexique* (1995) et le numéro 106 de *Langue française* (1995). Ces ouvrages

¹ L'*exemple* utilisé comme hyperonyme comprend à la fois la citation littéraire et l'exemple forgé. Voir *infra*.

² « Ainsi se dévoile la nature de l'exemple : un énoncé où le mot-entrée signifie la chose nommée, et un énoncé qui signifie le mot-entrée comme signe-nommant » (Rey-Debove 1971 : p. 261).

³ Pierre Gilbert a aussi signé un article paru dans les *Cahiers de lexicologie*.

contribuent tous à leur manière à mettre en exergue les thématiques privilégiées des citations, les filiations entre les dictionnaires, le rapport entre syntagme et citation, les traductions d'exemples signés dans les dictionnaires bilingues, etc.

2. RAPPEL HISTORIQUE

Avant d'entrer dans le vif de notre sujet, abordons en premier lieu les origines de la pratique citationnelle. La création de la citation remonte à Papias, figure importante du XI^e siècle en linguistique. Ce linguiste innove également en indiquant le genre des noms par des lettres et en introduisant les conjugaisons des verbes (Boulanger 2003 : 277). On retrouve déjà plus tôt au Moyen-Âge la citation propre aux dictionnaires mais aussi celle que l'on retrouve dans les florilèges d'*auctoritates*, qui sont des compilations de passages ou d'extraits pris aux auteurs classiques, aux Pères de l'Église ou aux philosophes (Boulanger 2003 : 393). En latin classique *auctoritas* dénommait les personnes et magistrats et renvoyait de manière générale à l'autorité des auteurs classiques, à des « modèles à imiter », extraits d'un ensemble, selon le sens du mot latin *exemplum* (Rey 1995b : 96). C'est par métonymie qu'il en est venu à désigner le texte puis l'extrait du texte. Dans une savante explication de l'histoire de la lexicographie, Jean-Claude Boulanger explique qu'au Moyen-Âge existaient quatre types de citations : 1. « on citait un extrait de texte que l'on faisait suivre ou précéder d'une référence plus ou moins précise ; 2. on citait seulement le nom de l'auteur ; 3. on citait le nom de l'auteur ainsi que le titre de l'œuvre ; 4. on citait le nom de l'auteur qui était lui-même cité par un autre auteur. » (Boulanger 2003 : 394). L'autorité constitue l'énonciateur de référence, reconnu et légitimé par le dictionnaire. Ainsi, citer « implique une littérarité et une fidélité des propos rapportés, ainsi qu'une attribution "glorieuse" des dits propos » (Rosier 2005 : 67) comme une allégation. En relayant l'acception juridique utilisée lorsque l'on cite en justice, le champ littéraire et linguistique a pratiqué, comme en matière juridique, la métonymie, en utilisant l'acception issue du latin *citatio* : « ce qui est appelé, invoqué ». En droit, on peut en effet employer *citation*, en désignant l'assignation à comparaître comme le document juridique en tant que tel. Mais il est vrai que la frontière est parfois tenue entre le champ juridique et celui de la littérature si l'on en juge par cette citation du *Grand Robert électronique* (désormais *GRE*) : *Ce bel exploit (→ Conduite) lui a valu une citation.*

3. LA CITATION ENTRE LINGUISTIQUE, LITTÉRATURE ET ANALYSE DU DISCOURS

Mais c'est surtout par les champs littéraire et linguistique que l'emploi de *citation* est répandu. La phraséologie moins figée que celle de la doctrine juridique laisse place à des emplois plus variés. Ainsi *on bigarre, on bourre* ou *on entrelarde un discours de citations, on charge* ou *on gonfle un ouvrage de citations, on entasse des citations, on émaille un texte de citations* ou, au contraire, *on détache d'un texte une citation, on dépouille une citation de son contexte* ou *l'on estropie une citation* ou enfin, verdict final, *on écarte une citation d'un dictionnaire*. Tous ces exemples forgés provenant du *GRE* montrent l'activité d'extraction et d'inclusion de la citation que pratique tout citateur, notamment le lexicographe. Sa pratique ressemble en plusieurs points à celle pratiquée par le littéraire. Dans un premier cas, la citation est un objet prélevé d'un contexte linguistique, elle sert souvent à la

compilation d'anthologies de citations. Les connaître relève alors du signe distinctif d'un « honnête homme » et du discours scolaire (Aron 2004 : 99). Le deuxième contexte citationnel de la citation est celui où un citateur cite un auteur ou rapporte un discours tiers au sein d'un texte. Dans ce cadre, les théoriciens de l'analyse du discours privilégient l'appellation *discours direct*⁴ là où les sociologues et spécialistes de la communication emploient davantage le terme *citation* (Rosier 2002 : 28). Mais si la citation dans le discours linguistique se fait remarquer par son caractère polémique, critique, pédagogique, scientifique ou moralisateur, ses vertus sont tout autres dans la pratique lexicographique, comme nous le verrons, car dans ce contexte, elle a comme fonction de poser avant tout une équivalence sémantique avec le mot-entrée en plus de servir à l'élaboration et la sélection de la nomenclature (passant du statut d'occurrence à celui de citation).

4. LA RÉFÉRENCIATION ET L'ÉVOLUTION DE LA PRATIQUE CITATIONNELLE

Bien que la pratique citationnelle soit, comme nous l'avons mentionné, très ancienne, l'usage de l'informatique en lexicographie, plus récent, va en modifier de manière importante la pratique. C'est le *Trésor de la langue française* (désormais *TLF*) qui, le premier, a intégré l'informatique dans la technique dictionnaire de langue française. Grâce au Gamma 60 Bull, une importante documentation a pu être obtenue mécaniquement (en 1968, Matoré évaluait à 250 millions de mots tirés de textes littéraires). L'usage de l'informatique va résoudre un problème relatif au traitement des *corpus*, bien cerné par Matoré (1968 : 248) « Avant l'adoption des procédés mécanographiques, il était impossible, pour des raisons évidentes, de réunir assez d'exemples pour alimenter toutes les acceptions : certaines, par malchance, n'étaient pas illustrées par des citations ; d'autres, par contre, trop abondamment pourvues, étaient éliminées par l'auteur, ou bien elles restaient en surnombre (pratique fréquente chez Littré). »

L'usage de *corpus* lexicaux répond notamment à ce besoin de vérifier des emplois afin de dégager des éléments de phraséologie (collocation, allusions littéraires, citations, cooccurrences, syntagmes et expressions). L'élaboration et la consultation de *corpus* informatisés, jumelées au dépouillement systématique de sources récentes, comme les recommandations des commissions de terminologie et la presse quotidienne, permettent la revitalisation de la nomenclature, tâche qui s'avérait impossible avec les moyens techniques de la dictionnaire d'avant les années soixante-dix, quatre-vingt. Ainsi en va-t-il de l'amélioration de la technique lexicographique, mais aussi de l'accès plus facile aux œuvres des auteurs de la Francophonie.

Mais l'importante documentation révéla aussi des problèmes également rencontrés dans d'autres dictionnaires de grande envergure, entre autres celui de la référencement. Ainsi, faute de temps et d'espace typographique, ce sont souvent les datations qui se sont avérées imprécises, parce que l'on ne citait pas l'édition d'origine, ou l'on citait l'édition d'un jour-

⁴ Il s'agit même parfois d'autocitation comme *j'ai dit que* ou de citation dans lequel l'effacement de l'auteur se fait afin de valoriser la parole du groupe. L'hétérocitation ainsi que nous pourrions la nommer est ainsi un processus d'anonymisation.

nal intime couvrant plusieurs décennies. Depuis peu, le laboratoire d'analyse et traitement informatique de la langue française (ATILF) procède ainsi à une révision progressive de la rubrique « Étymologie et histoire » de l'article lexicographique du *TLF* en y réinjectant le résultat des recherches récentes et des sources nouvellement publiées.

Quant au *Grand Robert* (désormais *GR*), les indications isolées données à la fin des citations sont en fait des références bibliographiques incomplètes, qui n'ont qu'une fonction testimoniale, comme le mentionne Laurent Bray, et ne sont pas opératoires (1995 : 92). De plus, comme dans le cas du *DAALF*, le passage d'un grand ouvrage à un ouvrage plus petit implique une transformation des *corpus* de citations en fragments courts référencés, de quoi s'interroger sur la fidélité des sources en matière lexicographique.

5. PHYLOGENÈSE ET ONTOGENÈSE DANS LA PRATIQUE LEXICOGRAPHIQUE

À partir de cet exemple récent de filiation entre deux dictionnaires, c'est-à-dire le passage du *DAAL* (dont le dernier fascicule *Vénus à Zymotique* paraît en août 1964) au *Petit Robert* (première édition : 1967) (désormais *PR*), notre analyse diachronique du *corpus* dictionnaire permet d'observer deux types d'évolution de la pratique citationnelle. Pour les expliquer nous empruntons à la biologie animale les concepts de *phylogenèse* et d'*ontogenèse*. La *phylogenèse* du grec *phulon* :/espèce/et *genesis* :/engendrer/, signifie, en s'appliquant à la métalxicographie, la formation et les enchaînements des filières citationnelles au fil du temps entre dictionnaires de traditions différentes⁵. La *phylogenèse* dictionnaire est un phénomène ainsi connu de la tradition lexicographique, comme en atteste Sidney Landau : « *No one would dare risk overlooking an important sense because of having failed to check other dictionaries. However, citation files must supply new terms and new senses unless the new dictionary is to rely entirely on existing works. Citations grant the work an independent validity but are vastly more time consuming to collect and use than other sources. The mix of old and new material is therefore of crucial importance in determining the dictionary's schedule.* » (Landau 1989 : 228)⁶ Ainsi, le Hatzfeld et Darmesteter, qui paraît à partir de 1890, reprend les citations de Littré dont le *Dictionnaire de la langue française* paraît à partir de 1863.

Quant à l'*ontogenèse*, du grec *ontos* :/être/et *genesis*, elle signifie le développement d'une filière descriptive à l'intérieur d'une même tradition dictionnaire d'un dictionnaire source vers un ou des dictionnaires cibles. Les maisons Robert et Larousse nous fournissent à eux seuls des exemples convaincants. Ainsi les exemples non référencés dans le *Micro Robert* procèdent d'exemples référencés du *PR* (cf. Rey 1995b : 105). Aussi, fait remarquer A. Rey, de Pierre Larousse à Claude Augé (*Nouveau Larousse illustré*, 1897), puis à Paul Augé (*Larousse du XX^e siècle*, 1933), « le rôle de la citation littéraire s'amenuise, au profit

⁵ Alain Rey signale par exemple « En France, Landais, Poitevin, Bescherelle, puis Pierre Larousse illustrent cette tendance : aucun ne se passe d'extraits littéraires, continuant par là une politique mise en œuvre au XVIII^e siècle par la série des dictionnaires de Trévoux, qui l'avait hérité de Furetière. » (Rey 1995a : p. 20). Laurent Bray signale les erreurs d'interprétations du *Grand Robert* de citations provenant du Littré (1995 : p. 99).

⁶ Voir aussi p. 152. Voir aussi Rey, 1977 : p. 142, Matoré, 1968 : p. 221, Françoise Martin-Berthet, 1995 : p. 61.

de l'exemple anonyme, qu'il soit fonctionnel ou encyclopédique. De Paul Augé à Claude Dubois (*Grand Dictionnaire encyclopédique Larousse* [1982-1985]) c'est la citation littéraire, pas seulement la citation, qui disparaît. » (Rey, 1995a : 21) L'adage du Larousse puisé dans la correspondance de Voltaire (Lettre à Charles Pinot Duclos, secrétaire perpétuel de l'Académie française, 11/08/1760) « Un dictionnaire sans exemples est un squelette » perdant en partie son sens même si les illustrations peuvent remplir la fonction d'exemple. Pourtant le *GR* qui cite Voltaire en a retenu à l'entrée *squelette* et *citation* : *Un dictionnaire sans citation est un squelette* d'où cette possible confusion sur le fait qu'il n'y aurait ni exemples ni citations dans le *PL*.

6. LE RAPPORT ENTRE LE MOT-ENTRÉE ET LA CITATION

Le remaniement des citations est l'une des observations à laquelle le chercheur aboutit dans l'étude de l'ontogenèse lexicographique. Ce remaniement cherche toujours à obtenir un équilibre entre un espace typographique disponible et la qualité d'une information nécessaire pour décoder le sens d'un mot. Sur cette base, en s'inspirant de la théorie de l'information, nous pouvons aussi évoquer la *citation entropique* : qui nous renseigne peu sur le contenu du mot et la *citation informationnelle*, qui signifie à peu près tout⁷. Selon l'échantillon de Rey-Debove, le deuxième type serait plus fréquent. Pour le premier cas elle cite le *PR* : *Soubrette* : « Donner une dizaine de mille francs à la soubrette (Balzac) » et comme exemple du deuxième type à *Épuisette* elle donne parmi son échantillon un exemple tiré du *Dictionnaire du français contemporain* (désormais *DFC*) « Quand le pêcheur a amené la carpe au bord, il la retire de l'eau avec son épuisette. » Malgré la transparence du lien entre signifiant et son contexte citationnel, il n'en demeure pas moins que la contextualisation d'un item ne représente toujours qu'une « possible monstration d'une réalisation particulière parmi d'autres possibles ». (Corbin 2004 : 21)

7. FINALITÉ DE LA CITATION

Un autre aspect de la pratique citationnelle mérité d'être soulevé, c'est que la finalité de la citation change en fonction du projet dictionnaire. Selon les observations des linguistes Lionel Meney et John Sinclair, la finalité de la citation consistant à confirmer la définition, fournir un supplément d'information constituerait une vision théorique idéale de la citation. Dans son essai polémique où il justifie les principes lexicographiques de son dictionnaire, Meney explique que son dictionnaire est bivariétal dans la mesure où il a « *pour objectif la juxtaposition des termes qui distinguent deux variétés d'une même langue afin de faciliter le passage d'une variété à l'autre, autrement dit à des fins de traduction* ». Dans le cadre de ce projet dictionnaire, le but principal des exemples n'est pas de « *confirmer la définition* » d'un terme ni de « *fournir un supplément d'information* » [il reprend ici les propos des critiques] sur ce terme, mais de fournir des unités de traduction de L1 à L2 (Meney

⁷ Rey-Debove définit elle-même la citation comme un « signifié de forme prédicative "parlant d'un sujet" ». (1971 : p. 273). Il ne faut néanmoins pas négliger le fait que du point de vue de la différence entre le dictionnaire de langue et le dictionnaire encyclopédique, c'est plutôt le deuxième qui informe que le premier mais l'information relève de son approche encyclopédique alors que le premier fournit des articles et des citations pour définir en ayant évidemment recours à des « informations » sémantiques.

2002 : 37). Ainsi pourrait-on ouvrir le débat sur l'usage différencié de l'exemple ou de la citation dans les dictionnaires bilingues par rapport aux dictionnaires monolingues.

8. EXEMPLE SIGNÉ ET CRYPTOCITATION

La finalité et la fonction de la citation nous conduisent à commenter la nature même de la citation. Les terminologies ont évolué au fil des quarante dernières années, entre Bernard Quemada, Rey-Debove, Matoré, Rey, Robert Martin reprenant Rey-Debove, etc. : on atteste ainsi d'une part *énoncés-exemples* comprenant *exemples générés*, *exemples forgés* ou *construits*, *exemples non-référencés* (Rey 1995b) et *exemples observés* (Rey 1977), et d'autre part *citations référencées* (Rey, préface *Grand Robert*) comprenant *exemples signés* et *citations d'auteurs* (Alise Lehmann). On peut observer deux catégories d'exemples depuis que Jean-François Féraud en a fait la distinction explicite⁸ : les exemples signés, présentent un contexte d'usage d'un énoncé extra-dictionnaire, impliqué par la signature et les *cryptocitations* qui sont des exemples forgés par le lexicographe (des phrases ni vraies ni fausses (Martin 1989 : 600)). On retrouve dans la syntagmatique dictionnaire des exemples courts non signés (fragment de discours littéraire effacé) qui sont en fait le produit de l'observation de la distribution lexicale (étant souvent la réduction d'exemples cités comme dans le *TLF*). Nombreux auteurs ont affirmé que cette recontextualisation gommait en quelque sorte leur véritable nature en défigurant le contexte énonciatif d'origine⁹. Ainsi les unités de discours citées *faire son possible* (*Petit Larousse*, désormais *PL*) et *Ils se détendent* (*PR*) sont aussi présentes dans le *corpus* linguistique même s'il s'agit de cryptocitations (Rey 1995a : 23).

On peut également imaginer que le contexte exemplifié par la citation soit en décalage avec le savoir encyclopédique ou littéraire du lecteur qui permet de décoder le sens du contexte énonciatif. Il convient donc de se demander d'une part si la citation offre une garantie de stabilité suffisante pour la compréhension du contexte extralinguistique alors que son rôle dans la technique dictionnaire est renforcé par l'élargissement des *corpus*. Quelle est, d'autre part, l'utilité de l'élargissement de *corpus* pour le dictionnaire de langue générale à nomenclature plus réduite que des projets comme Le *TLF*, le *Grand Larousse de la langue française* (désormais *GLLF*) et le *GR*, sinon la recherche de néologismes et la variété des exemples (y compris dans les *corpus* oraux autrefois négligés par les lexicographes). Un projet dictionnaire comme celui du *Dictionnaire historique du français québécois* (désormais *DHFQ*) fonctionnant sur des principes étymologiques, historiques et philologiques, de même que le *TLF*¹⁰ et le *GR*, par l'étendue de leur nomenclature, impliquent le recours à

⁸ Voir Françoise Martin-Berthet, 1995 : p. 55.

⁹ Voir entre autres Martin, *op. cit.* : p. 603 ; Georges-Elia Sarfati cité par Pierre Corbin et Jean-Pierre Guillermin, 1995 : p. 13.

¹⁰ Le statut du *TLF* est relativement mitigé si l'on considère avec H. Caputo (1992 : p. 80) que « sa formule [est] "pseudo-historique" ; les attestations retenues couvrent la période 1760-1960. Ces deux dates sont trop éloignées pour définir un usage temporel effectif (une trentaine d'années) et trop rapprochées pour rendre compte d'un modèle culturel ».

un *corpus* dans une perspective d'attestation linguistique¹¹, de preuve de datation plutôt que de simple illustration d'un emploi.

Entre le dictionnaire philologique (*TLF* et *DHFQ*) et le dictionnaire de langue générale (*PR*, *Dictionnaire Hachette*), le recours aux sources est ainsi motivé par des raisons différentes. En outre, la nature même des sources citées peut varier. Cela nous amène à distinguer la *citation métalinguistique* de la *citation à contenu métalinguistique*. Un dictionnaire comme le *DHFQ* comporte, comme le *Dictionnaire historique de la langue française* (désormais *DHLF*), des exemples cités linguistiques (selon la terminologie de Martin (1989 : 600)), c'est-à-dire qui émanent du *corpus* littéraire, et des exemples cités métalinguistiques. Ces derniers impliquent la référence à une autorité issue d'un ouvrage de référence (dictionnaire, grammaire, etc.) bien que la référence soit ici indiquée au complet, contrairement au *PR*, et comporte des exemples forcément puisés dans plusieurs synchronies, étant donné sa vocation avant tout historique (ce qui le distingue du *Dictionnaire suisse romand* et du *Dictionnaire des régionalismes de France*).

L'usage d'exemples cités métalinguistiques peut porter à ambiguïté dans le cas où il s'agirait d'informations sur le signe, de prédication onomasiologique : « Cette alouette, que je nomme Lulu d'après son chant... » (Buffon, cité dans le *GR*). Elizabeth Grimaldi (2003 : 110) emploie ainsi le terme *jointure* pour désigner les « segments ou bribes de segments qui créent un lien syntaxique entre le mot et le discours définitionnel » tel qu'on le retrouve parfois à même les exemples cités métalinguistiques : *sous le nom de (sub voque incontenance)*, *exprimait cette vieille idée (sub voque collectivisme)*, *nommé (sub voque lion)*, etc. sont des prédicats et des tournures métalinguistiques parfois en position d'épithète ou d'apposition, qui correspondent à la fonction épilinguistique de Martin (1989 : 604)¹². Mais dans ce cas, même si la citation est métalinguistique, la source citée n'est pas proprement dite métalinguistique. Il faudrait donc différencier la *citation à contenu métalinguistique* (nature de la citation) et la *citation métalinguistique* (type d'ouvrage cité). Ces sources viennent souvent servir la rubrique étymologique sinon, plus souvent, celle de l'histoire du lexique, et ne peuvent être incluses que dans des dictionnaires ayant une portée philologique ce qui, dans des dictionnaires strictement de langue, comporte moins de pertinence (cf. Rey-Debove 1971 : 291). Ainsi, le commentaire géolinguistique implique pour un dictionnaire comme le *PR* la mention d'un auteur québécois ou d'autres pays de la francophonie que la France alors que le commentaire historique implique la plupart du temps la mention de dictionnaires appartenant à des synchronies différentes.

9. L'ARTICLE LEXICOGRAPHIQUE

Le passage en revue de la littérature métalexico-graphique nous permet d'arriver à différentes observations sur la citation, dont nous donnons quelques traits importants :

- la position dans le corps de l'article ;

¹¹ Ainsi dans le *DHFQ*, *bonnock* variante graphique de *banique*, « sorte de pain épais » est attesté chez Gabriel Roy dans *La rivière sans repos*, 1970, p. 210. Ici la citation d'un auteur fonctionne comme un *discours convoqué*.

¹² Voir Agnès Steuckardt, 2003 : p. 12.

- la fonction de la citation (rhétorique, idéologique¹³, etc.) ;
- longueur de la citation ;
- référence complète ou non en fin d’ouvrage : type de proprionymes (lexicographes, institution, auteurs antiques ou contemporains, toponymes, etc.¹⁴) ;
- exemple récent ou non ;
- pour quel type de mot ou d’emploi et pour quel motif choisit-on la cryptocitation au détriment de la citation (*cf.* Chantal Girardin 1995 : 32-33) ;
- quelle catégorie grammaticale est exemplifiée ;
- fréquence des auteurs cités et types d’auteurs cités (Richelet¹⁵, qui privilégie les bons auteurs et ceux de l’Académie) ;
- typographie (dans de nombreux dictionnaires comme le *Pluridictionnaire* et le *PR*, l’italique vaut pour l’exemple comme pour les syntagmes).

Tous ces facteurs interviennent dans le statut de la citation. Ainsi, « l’exemple peut servir de preuve en signifiant totalement un prédicat métalinguistique. Un groupe d’exemples peut servir de preuve en présentant pour chaque exemple des faits de langue constituant une classe fermée (variantes, formes conjuguées – singulier/pluriel) » (Rey Debove 1971 : 295).

10. POSITION DE LA CITATION DANS L’ARTICLE

Faute de pouvoir traiter de toutes les caractéristiques que je viens d’évoquer dans un texte de cette taille, je ne reprendrai que celle portant sur la position de l’article. La citation fait parfois figure de définition comme dans certains dictionnaires de langues de spécialité dans lesquels le texte de l’auteur, qui fait office de texte stéréotype d’usage, est repris dans son intégralité afin d’illustrer le sens du signe-chose en entrée. Citation et définition ne forment parfois qu’un seul fragment textuel montrant par là que le développement encyclopédique prime sur la définition de la chose nommée. D’autres fois, la lexicographie fait cohabiter sa définition avec l’exemple définitionnel. Selon Rey-Debove il s’agit là d’une pratique qui « permet de confirmer la définition par une autorité compétente, soit de discuter la définition en mentionnant des définitions un peu différentes pour le contenu. Cette discussion est [dès] lors plus terminologique que linguistique, et c’est surtout la chose qui est concernée (le découpage du monde). Il s’agit de s’entendre sur la nature de la chose (1971 : 286) ». Il

¹³ Dans ses répliques justifiant ses choix dans la pratique citationnelle du *Dictionnaire québécois-français*, Lionel Meney se défend d’accréditer une vision de la pratique citationnelle qui ferait du lexicographe un idéologue adhérent au message des citations et éliminant toute manifestation négative – ou prétendument telle – pour la société (2002 : p. 39). Voir Rey 1995b : 113.

¹⁴ Jean-Claude Boulanger observe que chez Richelet et Furetière, les noms propres et les formes propriales (Le Roy, palais, Parlemens, l’Elysée, Matignon), sont la plupart du temps localisés dans les exemples construits à l’exception des citations dans le Richelet et des développements encyclopédiques dans le Furetière.

¹⁵ Richelet serait, selon A. Lehmann, le « premier dictionnaire à recourir régulièrement à la citation. » (1995 : p. 39). Pour une étude statistique des auteurs du *GR* on se reportera à Pierre Gilbert. L’auteur observe dans le volume 6 du *GR* que les auteurs cités sont cités en moyenne 25, 15 dont Balzac et Hugo plus de 800 fois, mais aucun auteur belge ou québécois, et une préférence pour les auteurs du XIX^e siècle.

s'agit aussi de définition lorsque l'exemple forgé est placé en début de l'article lexicographique comme c'est le cas dans les dictionnaires d'apprentissage ou certains dictionnaires pour enfants¹⁶, afin de donner accès au sens.

Cela nous amène à constater que la position de la citation est un élément non négligeable sur le plan sémiotique. Ainsi, nous avons mentionné que placé au début de l'article, l'exemple forgé est surtout un exemple de longueur limitée, de structure syntaxique simple et sert d'accès au sens dans le cadre d'un dictionnaire bilingue, ou dans le cadre de la présentation d'un dictionnaire pour enfant à nomenclature réduite comme le *Robert pour enfants*. La citation située au milieu d'article est celle que l'on retrouve généralement dans les dictionnaires de langue. Enfin, l'exemple situé en fin d'article est le plus souvent une citation d'auteur. Dans bien des cas et ce jusqu'au début du XX^e siècle, la parémiologie était soit l'objet d'une sous-entrée, soit servait d'illustration d'une acception selon le choix structurel du lexicographe. L'inclusion des formes gnomiques (sentences, proverbes, adages) est une pratique héritée de la mode des maximes dans les salons élégants au XVIII^e et XIX^e siècle et est plus ou moins disparue de la technique lexicographique aujourd'hui, ne constituant qu'une *définition personnelle de la chose* parmi tant d'autres¹⁷. Il est en tous les cas pertinent de remarquer que l'article lexicographique et les recueils de citations ont ainsi des occurrences communes dans leur *corpus*, l'article lexicographique rangeant les citations que l'on retrouve dans les anthologies sous la marque *Allus. Littér. (PR)* ou *Prov.* Faisant office à la fois de sous-entrée ou de citation, le statut du fragment textuel variant tout en présentant des divergences de positionnement.

11. CONCLUSION

En dépit d'une analyse comparative annoncée en début d'article, notre propos, décidément trop pauvre en exemples, ne pouvait s'empêcher de retourner aux sources de la pratique citationnelle afin d'évoquer les parallèles possibles avec la pratique actuelle. Traditionnellement, la technique lexicographique s'appuyait sur l'érudition de certains lexicographes pour établir un *corpus* exemplaire d'auteurs permettant l'élaboration de citations et des types de distributions fréquents, ce qui permet aussi de consigner la syntagmatique en se basant sur la fréquence des syntagmes qui, faute d'espace, permettent de faire l'économie de la phrase complète. Mais la plus grande disponibilité des textes depuis le XIX^e siècle, et avec l'enrichissement de la langue au fil des siècles et des décennies, notamment grâce à l'industrialisation et la montée en flèche de l'activité néologique, rendent difficile la confection d'un dictionnaire en se basant sur la compétence linguistique d'une seule personne. De plus, le traitement de la documentation disponible se trouve facilité par les moyens informatiques comme jadis la linguistique historique avait révélé les variations des usages dans le temps, fortifiant ainsi la pratique étymologique et affinant forcément la pratique citationnelle.

¹⁶ Voir A. Rey, 1995b : p. 108.

¹⁷ Voir Josette Rey-Debove, 1971 : p. 288, Robert Martin, 1989 : p. 605.

BIBLIOGRAPHIE

- Aron (P.), 2004 : « Citation », dans Aron (P.), Saint-Jacques (D.) et Viala (A.), édit., *Le Dictionnaire du littéraire*, Paris, PUF, pp. 98-99, coll « Quadrige ».
- Boulangier (J.-C.), 2003 : *Les Inventeurs de dictionnaires, de l'éduba des scribes mésopotamiens au scriptorium des moines médiévaux*, Ottawa, Presses de l'Université d'Ottawa, XVII-547 pp.
- Collectif, 2004 : « Les proprionymes dans trois dictionnaires de l'époque classique », dans Mercier (L.) et Cajole-Laganière (H.), édit., *français du Canada-français de France VI, Actes du sixième colloque international d'Orford, du 26 au 29 septembre 2000*, Tübingen, Max Niemeyer Verlag, pp. 213-228.
- Bray (L.), 1995 : « Notes sur la référence bibliographique », dans *Lexique*, 12-13, pp. 91-103.
- Caputo (H.), 1992 : « Norme(s), langue(s) et dictionnaires monolingues », dans *Cahiers du CIEL, Facteurs d'hétérogénéité dans le lexique*, Paris, Centre interlangue d'études en lexicologie, Université Paris 7, pp. 73-92.
- Collectif, 1995b : « Du discours au discours par l'usage : pour une problématique de l'exemple », *Langue française*, n° 106, pp. 95-120
- Collectif, 1977 : *Le lexique, images et modèles, du dictionnaire à la lexicologie*, Paris, Librairie Armand Colin.
- Collectif, 1992 : « La presse et les modalités du discours rapporté : l'effet d'hyperréalisme du discours direct surmarqué », *L'Information grammaticale*, n° 94, pp. 27-32.
- Corbin (P.), 2004 : « Théorie et pratique de la formation professionnelle de lexicographes : treize ans d'Expérience Lilloise », dans *Proceedings of the Eleventh Euralex international Congress, Euralex 2004, Lorient, France, July 6-10, 2004, vol. I*, Université Bretagne Sud, pp. 13-35.
- Gilbert (P.), 1966 : « Les citations d'auteurs dans le *Dictionnaire analogique et alphabétique de la langue française* de Paul Robert », dans *Cahiers de lexicologie*, 9, pp. 113-121.
- Girardin (C.), mai 1995 : « Une doctrine jésuite de l'exemple *Le Dictionnaire royal augmenté* de François-Antoine Pomey », dans *Langue française*, 106, pp. 21-34.
- Grimaldi (E.), 2003 : « Glose et effets de glose dans les *Mémoires d'outre-tombe* et autres textes de Chateaubriand », dans A. Steuckardt (A.) et Niklas-Salminen (A.), édit., *Le Mot et sa glose*, Aix-en-Provence, Publications de l'Université de Provence, pp. [105]-127.
- Landau (S.), 1989: *Dictionaries : The Art and Craft of Lexicography*, Cambridge et [al.], Cambridge University Press, 370 pp.
- Lehmann (A.), 1995 : « La citation d'auteurs dans les dictionnaires de la fin du XVII^e siècle (Richelet et Furetière) », *Langue française*, n° 106, pp. 35-54.
- Martin (R.), 1989 : « L'exemple lexicographique dans le dictionnaire monolingue », dans Hausmann, (F.) (J.) et al., édit., *Dictionnaires, Encyclopédie internationale de lexicographie*, Berlin [et] New York, De Gruyter, pp. 599-607.
- Martin-Berthet (F.), 1995 : « Les citations "anonymes" dans le *Dictionnaire critique* de Féraud (1787) », *Langue Française*, n° 106, pp. 55-67.
- Matore (G.), 1968 : *Histoire des dictionnaires français*, Paris, Larousse.
- Meney (L.), 2002 : *Polémique à propos du Dictionnaire québécois-français*, Montréal, Guérin éditeur.
- Quemada (B.), 1968 : *Les dictionnaires du français moderne (1539-1863). Étude sur leur histoire, leurs types et leurs méthodes*, Paris, Didier.
- Rey (A.), 1995a : « Le statut du discours littéraire en lexicographie », *Lexique*, 12-13, pp. 17-32.
- Rey-Debove (J.), 1971 : *Étude linguistique et sémiotique des dictionnaires français contemporains*, Paris [et] The Hague, Mouton, 329 pp.
- Rosier (L.), 2004 : « La circulation des discours à la lumière de l'effacement énonciatif : l'exemple du discours puriste sur la langue », dans *Langages*, n° 156, pp. 65-78.
- Steuckardt (A.), 2003 : « Présentation », dans Steuckardt (A.) et Niklas-Salminen (A.), édit., *Le Mot et la Glose*, Aix-en-Provence, Publications de l'Université de Provence, pp. [5]-17.

CORPUS DE DICTIONNAIRES

- Dubois (J.), *et al.*, 1966 : *Dictionnaire du français contemporain*, Paris, Librairie Larousse.
- [EN COLL.], 1905, *Petit Larousse illustré*, Paris, Larousse (tirages ultérieurs aussi consultés).
- Guilbert (L.), Lagane (R.) et Niobey (G.) *et al.*, 1971 : *Grand Larousse de la langue française*, en six volumes, Paris, Librairie Larousse.
- Hatzfeld (A.) et Darmesteter (A.) avec le concours de Thomas (A.), 1964 : *Dictionnaire général de la langue française du commencement du XVII^e siècle jusqu'à nos jours, précédé d'un traité de formation de la langue [...]*, réimpression intégrale, Paris, Delagrave. [L'édition d'origine date de 1889-1901].
- Knecht (P.) et Thibault (A.), 1997 : *Dictionnaire du suisse-romand*, Genève, Zoé.
- Littre (É.), 1863-1873, 1877 : *Dictionnaire de la langue française...*, Paris-Londres, Hachette, [Réimpression : 1873-1883, 4 vol. et 1 vol. : *Supplément*, 1883].
- Meney (L.), 1999 : *Dictionnaire québécois français*, Montréal, Guérin.
- Micro Robert Dictionnaire du français primordial*, Paris, Le Robert, 1983, 2 tomes.
- Pluridictionnaire Larousse, Dictionnaire encyclopédique de l'enseignement*, Paris, Librairie Larousse, 1975.
- Poirier (C.), édit., 1998 : *Dictionnaire historique du français québécois ; Monographies lexicographiques de québécoisismes*, Sainte-Foy, Presses de l'Université Laval.
- Robert (P.), 1951-1964 : *Dictionnaire alphabétique et analogique de la langue française. Les Mots et les Associations d'idées*, Paris, Société du Nouveau Littre, 1960. [Paru sous le titre *Grand Robert de la langue française*, 2^e édition, Paris, Le Robert, 1985 et une version remaniée et amplifiée en 2001].
- Rey-Debove (J.) et Rey (A.) (texte remanié et amplifié sous la dir. de), 1993 : *le Nouveau Petit Robert, dictionnaire alphabétique et analogique de la langue française*, Paris, Le Robert [3^e édition, édit., or. 1967, 2^e édit., 1977].
- Rey, (A.), édit., 1992 : *Dictionnaire historique de la langue française*, Paris, Le Robert.[2 édit., 1993].
- Rezeau (P.), édit., 2003 : *Dictionnaire des régionalismes de France*, Bruxelles, Deboeck.

ANALYSE SÉMANTICO-DISCURSIVE DES COLLOCATIONS LEXICALES EN *CORPUS* SPÉCIALISÉ : LA BASE *CONNAISSANCE-S*

Estelle Dubreil et Béatrice Daille

LINA FRE CNRS 2729

Université de Nantes

Nantes, France

Résumé : *L'association terminologique préexistant au phénomène collocationnel est qualifiée d'arbitraire et rapportée de ce fait à une question d'usage linguistique conventionnel. Toutefois, en soumettant les collocations à une analyse sémantico-discursive, il s'avère que cette association terminologique est parfois prévisible, puisque sémantiquement calculable sur la base des possibles argumentatifs présents dans la signification lexicale de la base, ce qui ouvre une perspective de description possible de leur principe de composition en termes de blocs d'argumentation.*

Mots-clés : *collocation épistémique, corpus spécialisé, linguistique de corpus, sémantique lexicale*

1. PROBLÉMATIQUE

Le discours scientifique spécialisé (DSS) se présente comme le discours de la connaissance. Il tend à informer et est par conséquent considéré comme étant objectif (Mortureux 1985 ; Beaune 1988). Toutefois, ce postulat va à l'encontre d'un fondement théorique propre à la Sémantique des possibles argumentatifs – SPA – (Galatanu 2000, 2004, 2005), puisque ce modèle, à l'interface entre la sémantique théorique et l'analyse du discours, considère le langage général comme étant par essence évaluatif, c'est-à-dire porteur de visées subjectives par nature. Or, bien que singulier par ses préférences lexico-syntaxiques, le DSS recourt nécessairement au langage général pour se construire. Aussi, le point de départ de cette recherche consiste à tester l'hypothèse très générale selon laquelle le DSS est empreint de subjectivité, et le cas échéant de relever et décrire les mécanismes sémantico-discursifs impliqués dans ce processus. Pour ce faire, nous avons constitué un *corpus* électronique spécialisé. L'aspect informatisé de notre *corpus* et le recours nécessaire aux techniques d'analyses liées à la linguistique de *corpus*, nous ont conduites à observer les phénomènes collocationnels. Traditionnellement, l'association terminologique préexistant aux phénomènes collocationnels est qualifiée d'arbitraire et rapportée de ce fait à une question d'usage linguistique conventionnel. Pourtant, suite au dépouillement des données, il nous a paru nécessaire de rapprocher cet objet d'un autre concept fondateur en SPA, le possible argumentatif. En effet, en soumettant les collocations à une analyse sémantico-discursive, il s'avère que cette association terminologique est parfois légitime, prévisible puisque sémantiquement calculable ; ce qui remet en cause leur caractère arbitraire et ouvre une perspec-

tive d'explication possible de leurs principes de composition en termes de blocs d'argumentation.

Après avoir défini ce que nous entendons par collocation et présenté le cadre théorique sémantico-discursif retenu pour cette étude, nous expliciterons l'intuition sous-jacente au rapprochement des deux concepts mobilisés. Nous détaillerons ensuite nos hypothèses de travail et exposerons la méthodologie générée pour les tester. À titre d'illustration, nous procéderons à l'analyse des collocations incluant *connaissance-s* et conclurons sur ces premiers résultats.

2. COLLOCATION

D'après Williams (2003), p. 5 : « Le terme "collocation" a été introduit dans les années trente par J. R. Firth, membre fondateur de l'école contextualiste britannique, pour caractériser certains phénomènes linguistiques de cooccurrence qui relèvent essentiellement de la compétence linguistique des locuteurs natifs (Firth 1957). »

Ce terme recouvre en fait des phénomènes linguistiques très variés relevant de la phonologie, pragmatique, sémantique, syntaxique et même statistique (van der Wouden 1997). La collocation est la plupart du temps définie comme une association syntagmatique restreinte trouvant sa place entre le syntagme libre et le syntagme figé. Par exemple, en français un *café noir* est un café sans adjonction de lait et si *café* conserve sa signification habituelle, l'adjectif *noir* prend un sens nouveau. D'ailleurs, en espagnol, un *café noir* se traduira par un *café solo*. Du fait de leur caractère idiomatique et donc du problème qu'elles posent à tous les apprenants d'une langue étrangère, elles ont été décrites et modélisées par de nombreux lexicographes : citons par exemple le *BBJ Dictionary of English Word Combination* (Benson *et al.* 1986), le *Dictionnaire explicatif et combinatoire* (Mel'čuk *et al.* 1984-1999). La collocation lexicale telle que définie par Haussmann (1997) est composée d'une base et d'un collocatif ; dans notre exemple ci-dessus, *café* est la base, *noir* est le collocatif. Elle possède un certain nombre de propriétés : récurrence, arbitraire et bonne formation syntaxique, qui ont été exploitées pour leur découverte en *corpus* (Daille et Williams 2001). La motivation sémantique des collocations selon Grossmann et Tutin (2003) repose soit sur des procédés sémantiques productifs, comme la métaphore, la comparaison ou la métonymie, soit sur des régularités sémantiques, par exemple *affreux* se voit associé un sens intensif en cooccurrence avec les noms de sentiments de polarité négative.

3. POSSIBLE ARGUMENTATIF

Le Possible argumentatif (PA) est un concept pivot dans la théorie SPA auquel il donne d'ailleurs son nom. Pour en mesurer la portée et l'enjeu, nous exposons brièvement les fondements du modèle.

La SPA se donne pour objectif de proposer une « description/représentation de la signification lexicale » (Galatanu 2004), basée simultanément sur des informations sémantiques (internes à la langue) et discursives (externes à la langue, donc contextuelles). La *significa-*

tion lexicale est entendue comme « une conceptualisation largement identique d'un individu à l'autre, [formant] une sorte de socle pour une intercompréhension réussie » (Kleiber 1999). Elle se présente donc comme un processus d'abstraction concrétisé par une unité lexicale et visant une stabilisation du monde dans la langue. Du point de vue de la SPA, la description de la signification lexicale doit pouvoir rendre compte d'une part, de la *dimension évaluative du langage*, par une représentation des potentiels argumentatifs des unités lexicales et de la portée axiologique associée, et d'autre part, de la *dimension dénotative* du langage général, par la représentation du monde tel qu'il est perçu et modélisé par et dans la langue. Dans cette optique, la SPA décrit la signification lexicale d'un mot comme étant composée de trois strates : un noyau, des stéréotypes et des possibles argumentatifs.

Le *noyau* est une entité sémantique, composée de traits nécessaires de catégorisation (TNC) et de prédicats abstraits. Les TNC rassemblent trois types d'informations : la nature grammaticale du mot, ses caractéristiques modales et son orientation axiologique. Les prédicats abstraits correspondent à des primitifs sémantiques invariants (Putnam 1994).

Les *stéréotypes* sont également des entités sémantiques. Ils correspondent à la déclinaison des prédicats abstraits du noyau par association avec d'autres invariants (représentations), dans des blocs de signification argumentatifs posés comme naturels (du type : cause/effet, symptôme/phénomènes, but/moyen...). On parle de *blocs d'argumentation interne* à la signification lexicale (Carel et Ducrot 1999).

Les *possibles argumentatifs* sont des entités discursives. Ils correspondent à l'association *prévisible* de deux unités lexicales liées par une relation d'argumentation et calculable sur la base des enchaînements linguistiques que la signification lexicale est susceptible de générer en contexte. Plus précisément, il s'agit de la concrétisation dans le discours de l'association du mot avec l'un des éléments de ses stéréotypes. On parle de *blocs d'argumentation externe* à la signification lexicale (Carel et Ducrot 1999), puisque consommés dans le discours. Un bloc d'argumentation externe représente donc l'association, dans une structure argumentative, du mot dont on construit la signification lexicale avec l'un des prédicats identifiés comme faisant partie de son argumentation interne.

Nous exemplifions ci-dessous les trois strates de la signification lexicale avec le substantif *colère* (Galatanu 2004).

Noyau : Traits nécessaires de catégorisation (TNC)

Prédicat nominal féminin singulier

Affectif et moral

Axiologique négatif

Noyau 1 : penser à faire mal à soi-même ou à autrui

Stéréotypes :

Donc penser mal

Donc mériter punition

Donc vouloir vengeance

Possibles argumentatifs : *colère intérieure...*

Noyau 2 : éprouver émotion-sentiment négatif violent

Stéréotypes :

Donc état psychique mauvais

Donc état physique mauvais

Donc agressivité

Donc agir violemment

Donc agir sans réfléchir

Possibles argumentatifs : *colère folle, violente colère...*

De par le lien argumentatif posé, les PA sont par essence évaluatifs. Ils s'organisent autour de deux faisceaux orientés respectivement vers l'un ou l'autre des pôles axiologiques, positif ou négatif.

[1] Elle est intelligente, donc douée à l'école.

Activation d'un PA axiologique positif, porté par la signification lexicale de l'adjectif *intelligente*.

[2] Il est lâche, donc inutile de compter sur lui en cas d'attaque.

Activation d'un PA axiologique négatif, porté par la signification lexicale de l'adjectif *lâche*.

Les PA sont des entités discursives soumises aux caprices du discours, ce qui signifie que le discours influe sur les potentiels argumentatifs axiologiques. D'après Galatanu (2005, à paraître), p. 11 : « Le noyau et les stéréotypes forment dans la SPA un dispositif de génération de séquences discursives argumentatives, « les Possibles argumentatifs », qui peuvent être activés dans les occurrences discursives, ou renforcées, voire interverties, par des phénomènes de contamination co- ou contextuelle. »

[3] Il était en colère, donc il a frappé son ami.

PA conforme à l'orientation axiologique des stéréotypes du substantif *colère*.

[4] Il était en colère, pourtant il a amorcé une réconciliation.

PA contraire à cette même orientation axiologique.

4. CONCRÉTISATION DE LA COLLOCATION PAR UN POSSIBLE ARGUMENTATIF

Partant de la définition de *collocation* au sens d'Hausmann (1997), entendue comme toute séquence textuelle constituée de deux lexèmes dont l'un conserve son sens habituel, syn-

taxiquement correcte et fréquente, et de la définition de *PA*, entendu comme l'association prévisible de deux unités lexicales liées par une relation d'argumentation et calculable sur la base des enchaînements linguistiques que la signification lexicale est susceptible de générer en contexte ; nous postulons qu'une collocation correspond potentiellement à la concrétisation d'un PA dans le discours. Pour vérifier cette hypothèse nous avons mené une réflexion duelle, mêlant compromis théoriques et ajustements méthodologiques.

Au niveau théorique, les compromis ont émergé de la confrontation des deux concepts mobilisés et se sont traduits par la formulation d'un double paradigme.

Soit la collocation repérée est la concrétisation d'un PA dans le discours.

[5] violente colère

Dans la description de la signification lexicale de *colère*, on trouve associé au noyau, entre autres le stéréotype *violent*. Aussi, la collocation *violente colère* (et inversement *colère violente*) est-elle *calculable* sur la base des PA existants. Dans ce cas, on parlera de *collocation argumentative* et nous vérifierons qu'elle est intrinsèquement axiologisée.

Soit la collocation ne correspond pas à la concrétisation d'un PA dans le discours.

[6] colère rouge

Dans ce cas, l'association terminologique n'est pas prévisible, on ne peut calculer (de façon évidente) *rouge* dans *colère*. On parlera alors de *collocation inédite*. En l'absence de relation argumentative directe, la collocation inédite ne devrait pas être axiologisée, ce qui contredit la dimension évaluative du langage. Aussi, nous formulons une nouvelle hypothèse concernant l'axiologisation des collocations inédites. Nous postulons qu'il y a un relais pris par le contexte, c'est-à-dire un processus évaluatif externe dû au discours.

[7] Il faut intervenir quand la personne entre dans une colère rouge, afin de faire cesser l'escalade de la violence. (requête Google, 28/09/05)

La présence dans le contexte d'« escalade de la violence », nous permet de conclure que la collocation *colère rouge* possède une orientation axiologique négative. Naturellement, nous n'excluons pas que la base puisse conserver son potentiel axiologique et le propager ainsi à toute la collocation.

5. MÉTHODOLOGIE

Les ajustements évoqués au niveau méthodologique pour tester nos hypothèses se sont traduits par la création d'un processus de validation pluridisciplinaire, mêlant linguistique de *corpus* et lexicographie.

5.1. Linguistique de *corpus*

Nous utilisons le *corpus* électronique spécialisé *TAL(N)*, représentatif de la communauté du traitement automatique du langage naturel (Dubreil, 2004) pour mettre à jour les collocations. Ce domaine de pratique scientifique est dynamique en France. Les communications sont nombreuses, essentiellement écrites, ce qui satisfait nos besoins en analyses textuelles

et offre une double perspective scientifique confortable. D'une part, le regroupement d'une quantité de données suffisante est possible, et d'autre part, cela laisse espérer une certaine diversité des auteurs, permettant de décupler les possibles subjectivités discursives et d'accéder à un panel lexical varié. En outre, l'exploitation légale des documents sous la forme numérisée n'a pas posé de problème ; les textes appartiennent à l'association loi 1901 du domaine, l'ATALA, et ont été délivrés sur autorisation. Le *corpus TAL(N)*, conçu en 2003, comporte environ 1 million d'occurrences sans la ponctuation. Il réunit 193 articles de recherche scientifique : 69 articles de revue *TAL* et 124 articles de conférence. Il est constitué de textes entiers exclusivement rédigés en français et relevant d'un même genre-registre dit *écrits scientifiques*. Tous les textes appartiennent à une même période consécutive d'une durée égale : la revue *TAL* 1998-2002, du vol. 39.1 au vol. 43.3 inclus et la conférence *TALN*, les actes de 1999 à 2003 inclus.

Ce *corpus* a fait l'objet d'un étiquetage morphosyntaxique avec *WinBrill* et d'une lemmatisation avec *FLEMM*. Toutefois, ces données ne sont pas nécessaires pour l'analyse sémantico-discursive qui nous occupe actuellement, aussi, la consultation du *corpus* se fait sur les fichiers au format .txt.

Nous rappelons ici que l'origine de ce travail prend elle-même la forme d'une hypothèse volontairement trop générale (Dubreil 2005) consistant à tester si oui ou non, le DSS est empreint de visées subjectivantes. C'est pourquoi il nous a semblé pertinent d'analyser prioritairement les unités lexicales liées aux modalités du jugement de vérité, c'est-à-dire les mots appartenant à la zone dite épistémique, liée au *SAVOIR*.

[8] savoir (substantif et verbe), connaître, connaissance (singulier et pluriel)

Du point de vue *linguistique de corpus*, le protocole de validation de nos hypothèses de recherche se décompose en deux étapes : le repérage des unités lexicales épistémiques et le dépouillement des données.

Le repérage des unités lexicales épistémiques

Savoir étant le mot emblématique de cette zone modale, nous avons débuté nos analyses par lui, en tenant compte de son ambivalence grammaticale. Nous avons ensuite adopté une double démarche ascendante et descendante. Premièrement, pour couvrir le plus objectivement possible la zone épistémique, nous avons utilisé le *Dictionnaire des synonymes* développé par le laboratoire de linguistique de l'université de Caen, le *CRISCO*¹, en quête des synonymes de *savoir*, que nous avons recherchés ensuite en *corpus*. La démarche descendante consiste à relever dans le *corpus* les mots appartenant à cette zone modale, sur la base de la liste de fréquence des mots qui le composent. Nous avons volontairement conservé, voire restitué pour chaque paradigme sémantique le substantif et le verbe associé. Outre le fait de pouvoir mener à terme une analyse contrastive si cela s'avérait fructueux, cette démarche permet présentement de regrouper plus facilement les collocations entre elles. Nous considérons par exemple que *connaissance lexicale* et *connaissance du lexique* représentent deux variantes d'une même collocation.

¹ Cahiers du CRISCO, n° 17, juillet 2004.

Le dépouillement des données

Le dépouillement des données se fait essentiellement à l'aide du logiciel *WordSmith Tools*² (Version 3.0). Il s'agit d'un logiciel d'analyse lexicale, développé par Mike Scott depuis 1996. Il intègre de ce fait, un grand nombre de fonctionnalités (fréquences, concordances, mots-clés...) permettant de faciliter l'analyse du comportement des mots et de leur utilisation au sein des textes. Sur la base des données recueillies, il convient d'établir une série de critères fiables et stables pour juger de la pertinence des données nécessitant une analyse. Nous avons donc épuré nos données selon quatre critères : statistique, pertinence, lexical et terminologique.

Critère statistique

Nous avons dû préciser en premier lieu, notre acception de *répétitif* présente dans la définition de collocation. De façon arbitraire, nous l'avons restreinte à :

- la cooccurrence de deux termes apparaissant au moins trois fois dans le *corpus* ;
- la cooccurrence de deux termes issus de trois articles différents ;
- la cooccurrence de deux termes employés par trois auteurs différents.

Ces trois critères devant être simultanément remplis, ils garantissent selon nous l'authenticité des candidats retenus, étant représentatifs des usages linguistiques de la communauté TAL.

Nous avons donc retiré tous les candidats peu fréquents.

[9] une connaissance détaillée ; la connaissance globale

Critères de pertinence

Conformément à nos hypothèses de travail, cette recherche porte sur les collocations lexicales, les seules à pouvoir former des blocs d'argumentation axiologisés. Nous excluons de travailler sur les collocations grammaticales, constituées d'un mot plein suivi d'une unité subordonnée (verbe + prép. : *s'abstenir de* ; nom + prép. : *sentiment envers* ; adj. + prép. : *absent de*).

Nous avons donc retiré tous les candidats composés de mots grammaticaux.

[10] connaissances de/du/des

Critère lexical

Nous avons retiré toutes les lexies. Nous entendons par lexies, les expressions figées faisant l'objet d'une entrée lexicographique spécifique à laquelle est associée une définition particulière. De fait, ces entités ne partageant pas la même signification lexicale que l'unité épistémique recherchée, elles ne peuvent se soumettre aux mêmes blocs d'argumentation.

[11] à ma/notre connaissance

² Site web, <http://www.lexically.net/wordsmith/index.html>

Critères lexicographiques

De même que pour le critère lexical, le critère lexicographique impose d'exclure les termes complexes associés à un domaine de pratique. Nous utilisons pour ce faire, deux dictionnaires en ligne, *EURODICAUTOM* et le *Grand Dictionnaire terminologique*³. Si le candidat considéré figure dans l'un ou l'autre des dictionnaires, il est exclu. En outre, nous recourons au besoin à la compétence linguistique des locuteurs de la communauté, qui se prêtent au jeu de l'expertise.

Nous avons donc retiré des termes comme :

[12] base (s) de connaissance ; connaissance du monde ; extraction de connaissances

5.2. Lexicographie

Concernant l'aspect lexicographique, pour la construction de la signification lexicale (SL), le protocole consiste à prendre les définitions d'un mot dans deux dictionnaires éventuellement de traditions différentes, mais d'objectifs communs et surtout comparables (description d'un même état de la langue). Nous ne retenons pour la construction du noyau et des stéréotypes que ce qui est commun aux différentes entrées. Nous utilisons pour cette étude le Lexis Larousse et le Grand Robert. La SL de *connaissance* varie selon que le terme est singulier ou pluriel.

Description de la SL de *connaissance* au singulier.

Noyau : Traits nécessaires de catégorisation (TNC)

Prédicat nominal féminin singulier

Épistémique

Axiologique positif

Noyau 1 : processus : acquisition de la représentation mentale

Stéréotypes :

Donc viser une compétence

Donc étudier par la pratique en vue de maîtriser une compétence

Possibles argumentatifs : *connaissance de l'anglais*

Noyau 2 : résultat : fait de savoir

Stéréotypes :

³ *EURODICAUTOM* : <http://europa.eu.int/eurodicautom/Controller>

Grand Dictionnaire terminologique : http://w3.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp

Donc impression, intuition, sensation...

Donc certitude, compréhension...

Donc expérience, pratique...

Possibles argumentatifs : *connaissance intuitive ; profonde connaissance ; connaissance abstraite*

Description de la SL de *connaissances* au pluriel.

Noyau : Traits nécessaires de catégorisation (TNC)

Prédicat nominal féminin pluriel

Épistémique

Axiologique positif

Noyau : résultat : représentation (s) mentale (s) acquise (s)

Séréotypes :

Donc posséder des savoirs

Donc approfondir des savoirs...

Possibles argumentatifs : *enrichir ses connaissances ; connaissances élémentaires ; ensemble de connaissances*

6. RÉSULTATS

Nous avons extrait les 723 occurrences du terme *connaissance-s* singulier et pluriel. Les collocations construites sur la base de *connaissances* au pluriel sont nettement plus nombreuses que les collocations construites sur *connaissance* au singulier, dans une proportion de 4 pour 28, indépendamment du nombre d'occurrences.

Les résultats pour *connaissance* au singulier nous ont montré qu'à une exception près, toutes les collocations étaient de nature argumentative.

Ex. : bonne connaissance, connaissance approfondie, connaissance particulière

Dans tous les cas, la collocation conserve son potentiel axiologique positif.

La seule collocation inédite trouvée concerne *source-s de connaissances*. Toutefois, le statut inédit de la collocation tient au fait que *connaissance* est employé au singulier. Au pluriel, la collocation devient argumentative. D'ailleurs, on retrouve cette collocation dans sa version plurielle, avec une fréquence nettement plus élevée, 28 occurrences, contre 4 occurrences dans sa version au singulier. Aussi, sommes-nous tentés de conclure que la collocation juste est *source-s de connaissances*.

Les résultats pour *connaissances* au pluriel nous ont appris que toutes les collocations, là encore à une exception près, étaient de nature argumentative.

Ex. : ensemble-s de connaissances, fondée sur des connaissances, utiliser des connaissances...

Là encore, dans tous les cas, la collocation conserve son potentiel axiologique positif.

La seule collocation inédite initialement trouvée concerne *acquisition de connaissances*. Le statut inédit de la collocation tient à l'incompatibilité entre le processus *acquisition* et le résultat *connaissances*. Toutefois, bien que fréquente, 15 occurrences, la collocation semble s'apparenter davantage à un terme, représentatif d'une technique associée à un domaine (acquisition de connaissances à partir de texte ; technique d'acquisition de connaissance).

7. CONCLUSION

La faible quantité de données observées pour cette présentation nous empêche de statuer définitivement sur la pertinence de toutes les hypothèses de recherche posées, mais nous encourage fortement à poursuivre le dépouillement du *corpus*. En effet, l'analyse menée sur *connaissance-s* prouve l'existence de collocations argumentatives épistémiques. Dans ce cas, la collocation est bien axiologisée, ici positivement, conformément à la polarité de *connaissance-s*. Ces données confortent notre précepte général quant à la subjectivité du discours scientifique spécialisé et à son conformisme axiologique. Toutefois, l'existence de collocation inédite reste hypothétique et avec elle la faculté de la base à conserver son potentiel axiologique et le principe d'axiologisation extrinsèque par relais contextuel.

De toute évidence, cette étude remet en cause le caractère considéré comme arbitraire des collocations et ouvre une perspective d'explication possible de leur composition en terme de bloc d'argumentation, par relation « sémantico-discursive » motivée.

BIBLIOGRAPHIE

- Beaune (J.-C.), 1988 : « La vulgarisation scientifique : l'ombre des techniques », dans *Vulgariser la science : le procès de l'ignorance*, sous la dir. de D. Jacobi, coll. « Milieux »,
- Benson (M.), Benson (E.), Ilson (R.), 1986 : *The BBI Dictionary of English Word Combinations*. Amsterdam, Benjamins, 2^e éd, 1997.
- Carel (M.) et Ducrot (O.), 1999 : « Le Problème du paradoxe dans une sémantique argumentative », dans *Langue française*, sept. 1999, pp. 6-26.
- Daille (B.) et Williams (G.), dir., 2001 : « Collocation : computational extraction, analysis and exploitation », dans *Workshop at the 39th Annual Meeting and 10th Conference of the European Chapter of ACL*.
- Dubreil (E.) 2004 : Comment assurer la représentativité discursive d'un *corpus* électronique spécialisé ? L'exemple du *corpus* TAL(N), communication orale, dans 1^{er} colloque jeunes chercheurs Col'Doc 2004 « La Construction des observables en sciences du langage », Paris X – Nanterre, 29-30 avril 2004.
- Dubreil (E.), 2005 : Contraintes de constitution d'un *corpus* spécialisé dédié à une analyse sémantico-discursive, dans *Actes 2^e colloque Jeunes Chercheurs Col'Doc 2005 « Recueil des données en Sciences du langage et constitution de corpus : données, méthodologie, outillage »*, Paris X – Nanterre, 16-17 juin 2005 [à paraître].

- Firth (J. R.), 1957 : *Papers in Linguistics 1934-1951*. Oxford, Oxford University Press.
- Galatanu (O.), 2000 : « Langue, discours et systèmes de valeurs », dans *Curiosités linguistiques*, Suomela-Salmi édit., Université de Turku.
- Galatanu (O.), 2004, « Sémantique des “possibles argumentatifs” et axiologisation discursive », dans Bouchardeau (I.), [s. d.] *Représentation du sens linguistique II*, Paris Louvain-la-Neuve, Duculot.
- Galatanu (O.), 2005 : « Sémantique et élaboration discursive des identités. “L’Europe de la connaissance” dans le discours académique », conférence plénière, dans *Actes de la conférence internationale-culturelle and cross-linguistic Perspectives on Academic Discourse*, May 2005, University of Turku, Finland, school of Classical and Romance Languages [à paraître].
- Grossmann (F.) et Tutin (A.), dir., 2003 : *Les Collocations : analyse et traitement*, coll. « Travaux de recherches en linguistique appliquée », Paris, Éditions De Werelt.
- Hausmann (F.-J.), 1997 : « Tout est idiomatique dans les langues », dans M. Martins-Baltar (édit.), *La Locution entre langues et usages*. Paris, ENS éditions, pp. 277-290.
- Kleiber (G.), 1999 : « Problèmes de sémantique », dans *La Polysémie en question*, Nancy, Presses universitaires du Septentrion.
- Mel’čuk (I.), 1984, 1988, 1992, 1999 : *Dictionnaire explicatif et combinatoire du français contemporain : recherches lexicosémantiques*. vol. I, II, III, IV. Montréal, Presses de l’Université de Montréal.
- Mortureux (F.), 1985 : « Linguistique et vulgarisation scientifique », dans *Popularization of science*.
- van der Wouden (T.), 1997 : *Negative contexts*, Routledge, London.
- Putnam (H.), 1994 : *Le réalisme à visage humain*, Paris, Armand Colin.
- Williams (G.), 1998 : « Collocational Networks : Interlocking Patterns of Lexis in a corpus of Plant Biology Research Articles », dans *International Journal of Corpus Linguistics*, 3 (1), 151-171.
- Williams (G.), 2001 : *Les Réseaux collocationnels dans la construction et l’exploitation d’un corpus dans le cadre d’une communauté de discours scientifique*, Lille : Presses universitaires de Septentrion.
- Williams (G.), 2003 : « Les collocations et l’école contextualiste britannique », dans Grossmann (F.) et Tutin (A.) édit., *Les Collocations : analyse et traitement*, coll. « Travaux de recherches en linguistique appliquée », Paris, Éditions De Werelt

LE NOM PROPRE EN CONTEXTE – UNE APPROCHE LEXICOLOGIQUE

Jean-Louis Vaxelaire
CLILLAC (CIEL)
Université Paris VII
Paris, France
jl.vaxelaire@wanadoo.fr

Résumé : Les études sur le nom propre occultent généralement la dimension contextuelle pour une raison simple : la sémantique référentielle qui domine les débats s'intéresse plus aux individus désignés par les noms propres qu'aux noms en eux-mêmes. Si l'on veut décrire correctement les noms propres, il devient évident que seule une approche lexicologique, basée sur des textes, est efficace.

Mots-clés : nom propre, contexte, lexicologie, sens

INTRODUCTION

Le nom propre a longtemps été un sujet marginal en linguistique diachronique. Si l'on excepte les opuscules¹ de Gardiner (1954), Zabeeh (1968) et Algeo (1973), il faut en effet attendre les années 1980 pour trouver des travaux conséquents sur ce thème. La syntaxe du nom propre en français est désormais bien détaillée, mais il faut tout de même avouer que la conception lexicologique et lexicographique des noms propres demeure souvent élémentaire : pour une majorité d'auteurs, les noms propres réfèrent à des individus du monde physique ou de la fiction et sont en conséquence étudiés de manière totalement décontextualisée.

Dans cette optique, puisque le nom propre est un désignateur rigide² pour reprendre l'expression de Kripke (1980), un nom qui désigne le même référent dans tous les mondes possibles, quel intérêt y aurait-il à analyser dans des textes (qui sont forcément variables) ce point fixe ?

L'aporie de cette thèse est que le nom propre n'est plus perçu comme étant un signe linguistique mais comme l'expression d'une réalité physique, certains logiciens refusant même l'appellation *nom propre* pour les noms de fiction que sont *Cerbère* ou *Sherlock*

¹ Malgré leur intérêt théorique certain, ces livres ne font respectivement que 77, 78 et 94 pages.

² L'exemple du changement de nom du département des Côtes-du-Nord en Côtes d'Armor en 1990 démontre que c'est le numéro qui est rigide (pour des raisons pratiques dans ce cas) et pas le nom.

*Holmes*³. Pour le linguiste, le nom propre est pourtant — et avant tout — une lexie. Les considérations d'ordre métaphysique ne sont pas de son ressort.

Nous verrons que c'est en abordant le nom propre comme n'importe quelle autre partie du discours, c'est-à-dire sans *a priori* ontologique particulier, et en l'observant dans les contextes où il apparaît, qu'il est enfin possible d'obtenir une analyse linguistique plus juste.

1. NOM PROPRE OU NOM COMMUN ?

La question de la classification d'un nom dans la catégorie des noms propres ou des noms communs ne semble poser aucun problème. D'ailleurs, on remarque que la plupart des auteurs qui écrivent sur le nom propre ne prennent pas la peine d'en donner une définition, comme si celle-ci était évidente. En dehors des exemples classiques que sont *Socrate* et *table*, il existe cependant une multitude de cas problématiques, dont certains sont liés au contexte. Pourtant, il faut avouer que même les linguistes attentifs aux contextes semblent les oublier dès qu'il est question de nom propre.

Une des raisons principales du traitement différent que subissent les noms propres découle de l'omniprésence des théories sémantiques issues de la logique. Ainsi, l'idée que le nom propre n'est qu'une étiquette vide de sens remonte au début du XVIII^e s. avec Reid (1983) et Mill (1988). À la suite de Mill, Funke (1925) affirme que le nom propre est similaire à la cote d'un livre dans une bibliothèque, il n'est donc rien de plus qu'une étiquette collée sur un individu. Cette conception fixiste du nom propre, issue de la philosophie du langage, est encore largement répandue en linguistique. Pour demeurer dans la philosophie du langage, prenons l'exemple de Dummett qui se demande si *Marxism*, *quantum mechanics* ou *February* sont des noms propres. Nous pensons à l'instar de Willems (2000 : 106) qu'il est impossible d'obtenir la réponse en étudiant ces mots de manière isolée.

Si l'on s'en tient à la définition classique du nom propre, il désigne un individu particulier. Est-ce que *février* et *marxisme* désignent des individus particuliers ? Les avis divergent nettement selon les chercheurs.

D'autres exemples peuvent s'avérer problématiques :

[1] « Comparée à la tante Clara, ma si frémissante maman est un véritable Gary Cooper » (Roth, *Portnoy et son complexe*, Paris, Gallimard, 1970, p. 77).

[2] « Les tablettes découvertes depuis 1928 livrent une extraordinaire documentation sur cette Suisse de l'Antiquité, marchande et pacifique. » (*L'Express*, n° 2781, 18-24 octobre 2004)

[3] « Il devint un des héros populaires de la Révolution. » (*Petit Robert des noms propres*, article « Marat »).

³ D'après Sainsbury (2001 : 181) et Lejewski (1957 : 251), les noms de la mythologie tels que *Pégase* ou *Cerbère* ne sont pas des noms propres car ils ne nomment rien et Kripke a annoncé dans ses *John Locke Lectures* de 1973 qu'ils n'étaient que des « pretend names » (Hilpinen 1995 : 302).

[4] « La cause célèbre *Brown contre le Ministère de l'Éducation* en 1954 » (Taylor, *Le malaise de la modernité*, Paris, Cerf, 1992, p. 120).

[5] « Depuis Vilnius, Barclay s'est comporté plus en ami qu'en maison de disques. » (*Télérama*, 07/09/05)

Dans l'exemple [1], est-ce que la mère du héros du roman s'appelle *Gary Cooper*? La réponse est évidemment négative. Alors que *Gary Cooper* est couramment considéré comme un nom propre, il ne désigne pas un individu spécifique dans cette citation puisque l'auteur crée une classe de véritables *Gary Cooper*. À partir du moment où une classe est constituée, il devient impossible de parler de nom propre dans le sens de la définition traditionnelle, qui indique qu'un nom propre renvoie à un individu précis. Il ne reste alors que deux options : soit on modifie la définition de référence du nom propre, soit on considère que Gary Cooper est dans ce contexte un nom commun désignant les individus ayant telle et telle propriété (*i.e. être viril, héroïque, etc.*).

On peut tirer des conclusions similaires de l'exemple suivant : on ne parle pas du pays dont la capitale est Berne mais d'un pays qui possède des qualités de la Suisse, en l'occurrence être marchande et pacifique. Quelques linguistes estiment à l'instar de Gary-Prieur (1996) qu'il n'est pas possible de parler d'antonomase avec ce type d'exemple puisqu'ils ne sont pas lexicalisés. On peut effectivement contester cette appellation mais on ne peut considérer que *Suisse* demeure un nom propre dans ce contexte puisqu'il cesse de désigner son référent initial⁴.

Le terme d'antonomase s'applique également dans le sens inverse, comme dans l'énoncé [3] : pour nommer un événement, on emploie souvent un nom commun qui devient un nom propre, ici *révolution*.

Le quatrième cas relève d'un registre différent. Il pose la question de la diversité de la classe des noms propres. On limite souvent cette classe aux anthroponymes et aux toponymes (certains ouvrages pourtant consacrés au « nom propre » comme l'indique leur titre limitent leur propos aux seuls noms de personnes et noms de lieux). Il y a pourtant de nombreux autres individus particuliers que sont, entre autres, les partis politiques, les bateaux, les événements ou dans cet exemple les noms de causes, de lois, etc.

Enfin, avec le dernier exemple, on s'aperçoit qu'un nom propre peut servir à désigner autre chose que son référent initial, un événement (le meurtre de Marie Trintignant à Vilnius) dans [5] ou une date lorsqu'on lit des énoncés tels que « Depuis Dante... » ou « Depuis Rimbaud... »⁵.

Le poids du contexte pour distinguer les noms propres des noms communs intervient également dans le cas des noms de marque et de produit. Nous estimons, à la suite de divers

⁴ *Suisse* a, par antonomase, désigné d'autres lieux dont le relief rappelait la Confédération helvétique. Toutefois, certains noms ont été fixés et sont devenus officiels (la Petite Suisse luxembourgeoise, la Suisse normande, la Suisse saxonne, la Suisse franconienne, etc.).

⁵ « Rambervillers, comme beaucoup d'intellectuels depuis Dante et la Renaissance vit dans un idéal supranational et chrétien » (<http://www.gazette-vicoise.com/structures/museegdlt7.htm>). « Depuis Rimbaud, l'art a fait le saut ou l'a indiqué. » (<http://e-crisis.hautetfort.com/>).

auteurs tels que Lerat (1995) ou Jonasson (1994), qu'il est nécessaire de bien différencier ces deux types de noms puisque leur statut respectif est bien distinct, bien que leur signifiant soit souvent identique. Ainsi, dans « Acheter une BMW », une *BMW* est un exemplaire de voiture (nom de produit), il s'agit donc d'un nom commun : on peut changer son déterminant, mettre le nom au pluriel, lui ajouter un adjectif, etc. À l'inverse, dans « BMW a annoncé un programme de rachat d'actions », *BMW* est le nom d'une marque unique.

Les critères qui servent à distinguer les noms propres des noms communs sont le plus souvent formels (majuscule, absence de déterminant, etc.) mais ils ne s'appliquent pas dans de nombreuses langues. Il suffit de prendre l'exemple du français pour s'apercevoir que l'absence de déterminant n'est pas suffisante ; on recense en effet de nombreux noms propres nécessitant un article (*la France, le Balto, le parti communiste*, etc.). De même, l'argument des majuscules ne fonctionne que dans quelques alphabets d'origine européenne et certainement pas dans les systèmes d'écriture arabes, chinois ou indiens. On a pu lire que les noms propres avaient des déclinaisons différentes des noms communs (en polonais par exemple). Toutefois, des langues comme l'albanais possèdent les mêmes déclinaisons pour les deux types de noms. D'après Roman (1999 : 47), en arabe, le nom propre est, comme le nom commun, créé sur une racine de trois consonnes et comporte également la modalité d'animéité. Il n'y a pas non plus un suffixe particulier au nom propre.

Le cas le plus intéressant pour nous est celui du chinois. Dans cette langue, la catégorie grammaticale dépend de la fonction, de la place dans la phrase, ce qui implique que, sans contexte, il n'y a pas de nom propre ou de nom commun.

La situation n'est pas foncièrement différente en français. Certains signes sont usuellement employés en tant que noms propres (*Jean, Venise*, etc.) mais, puisqu'ils peuvent être réutilisés en tant que noms communs (*les venises du nord*), c'est principalement le contexte qui permettra de distinguer la catégorie grammaticale du nom.

Enfin, il existe une dépendance contextuelle de la prononciation des noms propres : *Berger* doit-il être prononcé en tant que nom français, allemand ou anglais ? Seul le contexte permettra de sélectionner les éléments pertinents.

2. TRADUCTION

[6] « À Belfast, la rentrée des classes s'accompagne, pour les fillettes inscrites à l'école Sainte-Croix, d'une redoutable épreuve quotidienne » (*Courrier international*, 07/09/01)

Du point de vue de la traduction, contrairement à l'idée reçue qui voudrait que le nom propre soit intraduisible, le contexte (et à un niveau supérieur le genre textuel) dicte souvent le choix de modifier ou de conserver un nom propre dans la langue cible. Ainsi, on traduit souvent les noms dans la littérature enfantine (*le Petit Chaperon Rouge/Little Red Riding Hood* ; *le Chat botté/el Gato con botas*). De même dans des contextes didactiques ou interculturels, on donnera des traductions de noms de partis politiques ou d'associations. Dans l'exemple [6], le nom de l'école Holy Cross est modifié alors que ce type de nom est rarement traduit, probablement pour montrer qu'il s'agit d'une école catholique. Dans les travaux universitaires, certains noms propres sont uniquement employés pour leur valeur

culturelle et les traducteurs peuvent donc les remplacer par des références locales qui ne perturberont pas le lecteur : les *Socrates*, *Guy Fawkes* et *Harold Wilson* d'*A Short History of Linguistics* de Robins deviennent ainsi *Socrate*, *Ravaillac* et *Pompidou* dans la version française (*Brève histoire de la linguistique — De Platon à Chomsky*, Paris, Seuil, 1976, p. 82).

3. LEXICOGRAPHIE

Il nous semble primordial sur le plan lexicographique de redéfinir la place des noms propres dans les dictionnaires : ils doivent apparaître en tant que lexies et non plus seulement en tant qu'éléments culturels. Bien qu'ils soient publiés sous l'appellation *dictionnaire*, des ouvrages comme le *Petit Robert des noms propres* ou le *Hachette des noms propres* sont nettement plus proches des encyclopédies dans leur traitement des entrées. L'ordre alphabétique n'est ainsi pas toujours respecté (les différents *Martin* sont classés selon leur date de naissance et non par rapport à leur prénom ; un nom de personnage tel qu'*Arsène Lupin* est classé dans les A au lieu des L, etc.) et de nombreux noms communs apparaissent dans la nomenclature, ce qui est pour le moins surprenant dans un ouvrage consacré aux noms propres. D'un point de vue linguistique, le défaut principal de ces dictionnaires tient dans leur quasi-absence d'informations d'ordre syntaxique et de prononciations (alors que les événements ont montré que les francophones ne savent pas comment prononcer le nom de l'île thaïlandaise de *Phuket*).

L'exemple des antonomases est symptomatique du traitement lexicographique des noms propres et de leurs dérivés. Les seules antonomases présentes sont généralement très anciennes (*Tartuffe*, *Judas*). Il est tout de même incroyable que *Waterloo* comme synonyme de défaite soit une antonomase acceptée dans les dictionnaires italiens par exemple, mais pas dans les dictionnaires français. La situation des métonymies telles que *Washington* pour le gouvernement américain ou *Place Beauvau* pour le ministère de l'Intérieur français est identique puisqu'elles sont totalement absentes des dictionnaires.

La lexicographie peine également à traiter les expressions contenant un nom propre. Si l'on prend *Ce n'est pas le Pérou*, on s'aperçoit qu'elle est absente des dictionnaires de langue et des dictionnaires de noms propres puisque ceux-ci ne comprennent pas de phraséologismes. En lisant l'article *Pérou* d'un dictionnaire de noms propres, on apprendra que ce pays a vingt-trois millions d'habitants, que l'industrie minière est en déclin, etc., mais rien sur le sens de la locution. Le problème est similaire avec d'autres phraséologismes tels que *sortir de Saint-Cyr* ou *trouver son chemin de Damas*.

4. CONCLUSION

L'idée que le nom propre serait un désignateur rigide est particulièrement répandue en linguistique et ne semble pas pouvoir être contestée. Il s'agit pourtant d'une notion philoso-

phique difficilement adaptable dans d'autres disciplines⁶. Ce qui intéresse Kripke, c'est l'individu désigné par le nom mais pas le nom en lui-même.

L'héritage sémantique de la philosophie du langage trouble entièrement la perception du nom propre en linguistique. On peut lire que les noms propres n'ont pas de sens parce qu'ils désignent directement⁷, qu'ils n'ont aucun synonyme (*Washington* et *Maison Blanche* peuvent pourtant être interchangés dans certains contextes), etc. Toute une série de lieux communs qu'il faudrait prouver.

Dans un autre travail (Vaxelaire, 2005), j'ai écrit que les questions récurrentes dans la littérature sur le nom propre telles que *Quel est le sens de « Socrate » ?* ou *Quel est le sens d'« Aristote » ?* étaient mal posées. Il serait effectivement plus utile de demander : *Quel est le sens de « Socrate » dans ce texte ?* ou plus largement : *Quel est le sens de « Socrate » dans ce contexte ?* Tout sens est contextuel, il n'y a donc que des stabilisations *a posteriori* (dans les dictionnaires ou les grammaires), que ce soit pour les noms communs, les adjectifs, les verbes ou les noms propres. C'est uniquement en resituant les noms propres dans leurs contextes que l'on aboutit à une nouvelle perception, enfin linguistique, de cette catégorie.

BIBLIOGRAPHIE

- Algeo (J.), 1973 : *On Defining the Proper Name*, Gainesville, University of Florida Press.
- Funke (O.), 1925 : « Zur Definition des Begriffes "Eigennamen" » dans Keller (W.), édit., *Probleme der englischen Sprache und Kultur — Festschrift Johannes Hoops zum 60. Geburtstag überreicht von Freunden und Kollegen*, Heidelberg, Carl Winter's Universitätsbuchhandlung, pp. 72-79.
- Gardiner (A. H.), 1954 : *The Theory of Proper Names — A Controversial Essay*, Londres, Oxford University Press. [1940].
- Gary-Prieur (M.-N.), 1996 : « Les noms propres au figuré » dans Léonard (M.) et Nardout-Lafarge (E.) édit., *Le texte et le nom*, Montréal, XYZ, pp. 67-82.
- Hilpinen (R.), 1995 : « Peirce on Language and Reference » dans Ketner (K. L.) édit., *Peirce and Contemporary Thought — Philosophical Inquiries*, New York, Fordham University Press, pp. 272-303.
- Kripke (S.), 1980 : *La Logique des noms propres*, Paris, Minuit. [1972].
- Lejewski (C.), 1957 : « Proper Names », *The Aristotelian Society*, Supplementary, vol. XXXI, pp. 229-256.
- Mill (J.S.), 1988 : *Système de logique déductive et inductive*, t. I, Bruxelles, Pierre Mardaga. [1843].
- Pendlebury (M.), 1990 : « Why Proper Names Are Rigid Designators », *Philosophy and Phenomenological Research*, vol. L, n° 35, pp. 519-536.
- Reid (T.), 1983 : « Essays on the Intellectual Powers of Man », *Philosophical Works — With Notes and Supplementary Dissertations by Sir William Hamilton I/II*, Hildesheim, Georg Olms, pp. 213-508.

⁶ L'interdisciplinarité est souvent problématique. Ainsi, lorsque Pendlebury (1990 : 519) affirme que si les noms propres sont rigides, c'est parce qu'ils sont arbitraires, il oublie que les autres types de mots sont également arbitraires.

⁷ Lorsqu'on montre un objet à un enfant, on regarde d'abord cet enfant, il n'y a donc pas de désignation directe.

- Roman (A.), 1999 : *La Création lexicale en arabe — Ressources et Limites de la nomination dans une langue humaine naturelle*, Lyon, PUL.
- Sainsbury (M.), 2001 : *Logical Forms — An Introduction to Philosophical Logic*, Oxford, Blackwell.
- Vaxelaire (J.-L.), 2005 : *Les Noms propres—Une Analyse lexicologique et historique*, Paris, Honoré Champion.
- Willems (K.), 2000 : « Form, meaning, and Reference in Natural Language : A Phenomenological Account of Proper Names », *Onoma*, n° 35, pp. 85-119.
- Zabeeh (F.), 1968 : *What's in a Name ? — An Inquiry into the Semantics and Pragmatics of Proper Names*. La Haye, Martinus Nijhoff.

MOTS ET CONTEXTES :
ESSAI D'ANALYSE LEXICO-SÉMANTIQUE
DU VOCABULAIRE POLITIQUE AU CAMEROUN DE 1990 À 1992

Ladislav Nzessé

faculté des lettres et sciences humaines
département des lettres d'expression française
Université de Dschang
B.P 49, Dschang
République du Cameroun

***Résumé :** La présente communication se propose de faire une étude morpholexicale et sémantique du vocabulaire politique au Cameroun de 1990 à 1992. En 1990, avec le retour à la démocratie pluraliste, le Cameroun a connu une profonde mutation sociale et politique qui a eu une incidence particulière sur le marché linguistique. De nouveaux usages émergent dans le lexique du français, liés à la violence verbale dans le jeu politique et aux antagonismes idéologiques. À cet effet, pour accéder à la totalité sémantique de ces mots créés ou « relookés », il faut nécessairement intégrer l'extralinguistique, c'est-à-dire l'idéologique et le sociohistorique. Ce troisième axe constitue pour nous un élément fondamental pour une réinterprétation efficiente du signe linguistique.*

***Mots-clés :** décennie 1990, lexique, contenu idéologique, contexte de production, facteurs sociohistoriques, signe linguistique*

INTRODUCTION

La décennie 1990 marque un tournant décisif dans la vie sociopolitique des États africains. Le « vent nouveau » de la démocratie souffle jusque sur le pré carré africain, et l'un des défis que devra relever l'Afrique au cours de cette dernière décennie du XX^e siècle est la démocratisation de ses institutions. Aussi, au sortir du sommet de la Baule du 19 au 21 juin 1990, François Mitterrand, alors chef de l'État français, exigeait et encourageait la démocratisation de la vie politique en Afrique en ces termes : « l'aide de la France sera plus tiède envers les régimes autoritaires et plus enthousiaste envers ceux qui franchissent le pas vers la démocratie et le respect des droits de l'homme ». Au Cameroun, la démocratisation devient un impératif et une réalité effective avec la création de plusieurs partis politiques et associations de défense des droits de l'homme et des libertés. Le pays devient ainsi un véritable laboratoire dans lequel plusieurs conceptions de la gestion de la chose publique sont développées et débattues. Sur le plan strictement linguistique, c'est le lexique qui traduit les antagonismes politiques. Les acteurs politiques se servent de certains mots déjà existants dans la langue française et en créent d'autres pour traduire leurs options et opinions politiques. Les contenus de ces mots sont généralement différents selon le bord politique de leurs énonciateurs. Aussi, la valeur du mot n'est plus déterminée uniquement par les deux axes

classiques (sélection et combinaison) mais par un troisième, l'axe des paradigmes symboliques basés sur le pouvoir évocateur du signe linguistique. Le mot devient alors contenu idéologique et contexte de production. Ce contexte renvoie aux facteurs sociohistoriques. À cet effet, l'intégration de la perspective référentielle amène à reconsidérer la problématique saussurienne du signe linguistique.

Afin de cerner cette problématique, nous nous sommes limité à un ensemble restreint : il s'agit du vocabulaire politique au Cameroun de 1990 à 1992. Pour ce faire, nous aurons pour support deux organes de presse. L'on se proposera tout d'abord de définir les critères qui ont guidé le choix du *corpus* d'étude et de la période étudiée ; par la suite, l'on tâchera de faire une étude morpholexicale et sémantique des néologies, tout en montrant comment elles traduisent les oppositions politiques ; car ici le contenu du mot reflète à merveille la chapelle politique de l'énonciateur. Notre étude examinera plus particulièrement les néologies de forme et les néologies de sens. Ces deux catégories ont permis de rendre compte d'une réalité psychologique, sociopolitique et sociohistorique particulière.

1. CORPUS ET PÉRIODE D'ÉTUDE

Cette étude a pour support un *corpus* écrit. Il s'agit de deux organes de presse à savoir *Le Messenger* et *Challenge Hebdo*. Les raisons de leur choix¹ sont simples.

Le Messenger, il faut le dire, est le doyen de la presse privée au Cameroun. La date de naissance de cet hebdomadaire remonte à 1979. Sa ligne éditoriale est claire : neutralité à l'égard du pouvoir, défense des intérêts du peuple, fidélité et loyauté à l'égard de ce dernier. Nga Ndong (1989 : 14) le remarque en ces termes : « *Le Messenger* apparaît [...] comme un journal qui se veut à la fois loyal vis-à-vis du peuple, [...] sérieux et même investi comme les envoyés du Seigneur, d'une mission salvatrice ». De 1990 à 1992, au Cameroun, *Le Messenger* est bel et bien représentatif des joutes politiques avec la polémique d'alors. C'est un journal à vocation véritablement politique au point que même « ses articles sur le sport sont écrits sur un ton vif et polémique, destinés manifestement à en découdre avec l'adversaire » (Nga Ndong *op. cit.* : 26 – 27).

Challenge Hebdo, quant à lui, apparaît sur le paysage médiatique camerounais en 1991 avec le nouveau contexte politique qui a favorisé la création d'une multitude d'organes de presse. Le titre de ce journal est d'ores et déjà porteur de sa ligne éditoriale ou, si l'on veut, de sa personnalité : « *challenge* » qui veut dire « *défi, provocation* ». *Challenge Hebdo* refuse de rester spectateur des mutations sociopolitiques et choisit pour option de promouvoir la culture démocratique et de faciliter la participation de tous aux activités sociopolitiques de la cité. Le ton est au défi, à la polémique, aux débats contradictoires.

Comme on peut s'en rendre compte, ces deux journaux possèdent un trait commun à savoir les préoccupations d'ordre politique, et ont été le cadre d'expression par excellence des différents courants politiques présents au Cameroun de 1990 à 1992. D'ailleurs, la plupart des auteurs d'articles parus dans ces journaux étaient majoritairement des hommes politi-

¹ Dans le cadre de cette étude, nous avons dépouillé 26 numéros de *Le Messenger* et 24 numéros de *Challenge Hebdo*.

ques et des intellectuels de la société civile qui prenaient activement part au jeu politique dans le pays à cette époque.

Ce sont là les critères qui ont motivé le choix de ces deux organes de presse.

Quant à la période historique, nous l'avons voulu brève. Nous nous sommes limité à trois grandes dates de l'évolution politique au Cameroun à savoir 1990, 1991, 1992. Ce sont des moments historiques qui ont offert des modifications importantes de la structure sociale et politique au Cameroun :

- 1990 est l'année de l'ouverture démocratique matérialisée par l'abandon du monopartisme et le retour au multipartisme ;
- 1991 est l'année de la pratique du jeu politique dans tous ses contours, avec aussi un nouvel état de langue caractérisé par la violence verbale et les innovations terminologiques porteuses des réalités psychologiques et sociopolitiques du moment ;
- en 1992, le jeu politique se confirme avec, à la clé, des élections législatives et présidentielles qui apparaissent comme un test à la démocratie pluraliste, et les divers coups de force langagiers qui ont suivi la proclamation des résultats de ces élections.

Sur le plan purement linguistique, la délimitation de la période est motivée par le fait que les phénomènes lexicaux sont liés aux facteurs historiques, aussi doit-on replacer leur étude dans l'histoire, par le fait même que ces phénomènes dépendent immédiatement des causes extérieures à la langue (Dubois 1962 : 49).

2. LA CRÉATIVITÉ NÉOLOGIQUE

Le vocabulaire politique au Cameroun de 1990 à 1992 donne le plus souvent lieu à la néologie² de forme et à la néologie de sens. « La néologie de forme consiste à fabriquer [...] de nouvelles unités, alors que la néologie de sens consiste à employer un signifiant existant déjà dans la langue considérée, en lui conférant un contenu qu'il n'avait pas jusqu'alors » (Dubois *et alii* 1994 : 322).

2.1. La néologie de forme

Les néologismes de forme de notre champ ont lieu par composition et par dérivation.

D'une manière générale, la composition désigne la formation d'unités sémantiques complexes à partir d'unités lexicales susceptibles d'un emploi autonome. Les unités lexicales suivantes caractérisent notre champ d'étude :

- *Démocratie-éprouvette* : « démocratie » dont le principe essentiel est la protection des intérêts égoïstes des gouvernants au détriment de l'intérêt national.

² La néologie désigne le processus de formation de nouvelles unités lexicales.

[1] « C'est le peuple au nom de qui va de nouveau se jouer la tragi-comédie qui risque de payer de sa sueur et de son sang. Une fois de plus et au nom de la *démocratie-éprouvette* » (*Challenge Hebdo*, n° 38, 1991 : 3).

– *État-parti* : État qui devient tout simplement un parti politique. Au départ on a véritablement un parti politique puis le parti dévore l'administration. Et comme c'est cette administration qui produit les hommes de l'État, le parti absorbe l'État par ses hommes interposés, et l'État et le parti ne font plus qu'un.

– *Vote-sanction* : élection dont les résultats contestent la gestion des gouvernants.

[2] « La minorité qui a accepté d'aller aux élections a opéré un *vote-sanction* à l'endroit du pouvoir en lui refusant ses suffrages » (*Le Messenger*, n° 252, 1992 : 1).

[3] « Victime sans doute d'un *vote-sanction* comme plusieurs candidats du parti, Njoh Mouelle est assurément le symbole de l'essoufflement de la machine R.D.P.C³ » (*Le Messenger*, n° 252, 1992 : 10).

Les mots-valises constituent un type particulier de néologie par composition. Ces mots sont artificiellement formés en condensant deux parties des unités lexicales autonomes existant en français, pour former une nouvelle unité susceptible de jouir aussi d'une autonomie. Les exemples suivants sont éloquentes :

– *Démocrature* (Condense « démocratie » et « dictature ») : « régime politique » basé sur les principes de la dictature, mais se dit démocratique.

[4] « Allez-y ! Faites régner l'ordre et faites avancer la *démocrature* » (*Le Messenger*, n° 234-235, 1991 : 11).

– *Monocratie* (de « monolithisme » et de « démocratie ») : « régime politique » où la souveraineté est exercée par un groupuscule tout puissant. Dans ce type de « régime », le pluralisme politique et les débats contradictoires sont proscrits.

[5] « Cas des enseignants enrobés de force au service de la *monocratie* » (*Le Messenger*, n° 231, 1992 : 2).

– *Ventrocratie* (de « ventre » et de « démocratie ») : « Régime politique » où la souveraineté est exercée par une minorité animée par des intérêts égoïstes.

[6] « Cher ami votre *ventrocratie* résiste aux assauts de l'opposition » (*Le Messenger*, n° 230, 1991 : 2).

Quant à la dérivation, elle « consiste en l'agglutination d'éléments lexicaux, dont au moins un n'est pas susceptible d'emploi indépendant, en forme unique » (Dubois *et alii*, *op. cit.* : 163). Plusieurs occurrences caractérisent notre champ d'étude et le processus se réalise par suffixation :

³ Rassemblement démocratique du peuple camerounais. Parti au pouvoir au Cameroun depuis 1984.

– *Biyaïsme* (de Biya, nom du chef de l'État camerounais) : « doctrine politique » initiée et pratiquée par Biya. Au cœur de cette « doctrine » se trouvent la promotion des actions des enseignements et de la pensée du président Biya.

[7] « Ceux qui ont donc naturellement vocation à devenir des autorités morales préfèrent se cramponner aux petits avantages du *biyaïsme* » (*Challenge Hebdo*, n° 25, 1991 : 8).

– *Ahidjoïsation* (de Ahidjo, nom du premier chef de l'État camerounais après les indépendances) : dictature du président Ahidjo.

[8] « L'ahidjoïsation effrénée de la vie publique a empêché toute velléité démocratique au Cameroun après les indépendances » (*Challenge Hebdo*, n° 25, 1991 : 9).

– *Bétisation* (de Bété, tribu du chef de l'État camerounais) : présence dominante des Bétés dans tous les secteurs de la vie sociopolitique du pays.

[9] « Cette *bétisation* du pays a conduit à des frustrations énormes » (*Challenge Hebdo*, n° 38, 1991 : 4).

– *Présidaillon* (de président, valeur péjorative)

[10] « Or lorsque l'on sait que [...] les services de renseignement occidentaux sont présents sur les tropiques, faisant et défaisant nos *présidaillons*, il y a lieu de penser que plusieurs courent à leur perte » (*Le Messager*, n° 239, 1991 : 16).

– *Ministrion* (de ministre, valeur péjorative) : membre du gouvernement qui n'a ni de pouvoir ni de personnalité.

[11] « Vos ministrions qui arpentent les couloirs du palais sans jamais rencontrer le prince » (*Challenge Hebdo*, n° 25, 1991 : 10).

En contexte, tous ces mots composés et dérivés sont satiriques et sont contemporains aux réalités sociopolitiques. Il en va de même des néologismes de sens.

2.2. La néologie de sens

La néologie de sens constitue le plus vaste ensemble du vocabulaire politique au Cameroun de 1990 à 1992. À cet effet, ces mots sont devenus un « lieu d'ajustement et de glissements sémantiques » (Fuch 1984 : 49). Les exemples suivants en sont révélateurs :

– *Affameur de population* : membre du gouvernement spécialisé dans le détournement des fonds publics.

[12] « Les seuls interlocuteurs que méritent selon moi ces *affameurs de population*, [...] c'est le juge de tribunal correctionnel » (*Challenge Hebdo*, n° 23, 1991 : 4).

– *Ventriloque* : 1- individu dont la préoccupation principale est de manger, sans jamais réellement travailler ; 2- qui exploite à son profit les populations de plus en plus démunies.

[13] « Pendant ce temps, une suite de *ventriloques* ne cesse d'encenser la plèbe, et jouit même d'une immunité révoltante » (*Challenge Hebdo*, n° 26, 1991 : 4).

– *Tribu du ventre* : groupement d’individus homogènes au plan intellectuel, politique et social, ayant pour seule préoccupation la satisfaction de leurs intérêts égoïstes. Cette « tribu » regroupe une minorité composée des ressortissants de toutes les ethnies du Cameroun, et détentrice du pouvoir.

[14] « Kontchou ne représente pas les Bamilékés ; Mbouï n’incarne pas le peuple Bassa ; et Owona n’est pas mandaté par les Bétis pour défendre leurs intérêts. En revanche, ces trois hommes, et bien d’autres, sont originaires d’une seule et même ethnie : *la tribu du ventre* » (*Challenge Hebdo*, n° 3, 1991 : 11).

– *Politique du ventre* : manière d’exercer l’autorité, avec un souci exclusif de la satisfaction matérielle d’une minorité.

[15] « La fameuse commission indépendante sur les tristes événements de l’université où les étudiants sont finalement les boucs émissaires de la *politique du ventre...* » (*Challenge Hebdo*, n° 34, 1991 : 11).

– *Mont-Cameroun⁴ de la nullité politique* : le plus haut degré de la nullité politique.

[14] « Décidément, notre ministre de l’Information et de la Culture atteindra bientôt le *Mont-Cameroun de la nullité politique* » (*Challenge Hebdo*, n° 38, 1991 : 3).

– *Analphabète politique* : homme politique (en activité) qui ne lit pas ou très peu, n’écrit plus, ne conçoit plus, mais dicte et donne des instructions. À cet effet, il y a des gens pour lire et écrire à son compte.

[15] « Ces *analphabètes politiques* passent le plus clair de leur temps à signer notes, décisions, ordonnances [...] mais s’accaparent, de fait, aujourd’hui, le rôle de cultiver le jardin d’une démocratie tant relancée » (*Challenge Hebdo*, n° 38, 1991 : 11).

– *Pape des opposants politiques*. (Mongo Béti, écrivain camerounais aujourd’hui décédé) : le plus populaire et le plus téméraire des opposants politiques.

[16] « Mongo Béti triomphe à Douala [...] l’arrivée du *pape des opposants politiques* camerounais ne pouvait laisser personne indifférent. [...] L’homme reste fidèle à lui-même et les propos toujours acides » (*Challenge Hebdo*, n° 21, 1991 : 1).

– *Démocratie avancée* (dixit Paul Biya) : 1- Selon le parti au pouvoir, c’est une démocratie qui respecte véritablement les principes de la démocratie universelle ; 2- Selon l’opposition politique, c’est une « démocratie » qui se caractérise par des violations graves des principes élémentaires d’une démocratie vraie.

[19] « Le peuple remplira les urnes d’enveloppes [...]. C’est sa contribution à la *démocratie avancée* » (*Challenge Hebdo*, n° 48, 1991 : 2).

[20] « Les enlèvements, la torture et les violations des libertés du citoyen sont l’apanage de votre *démocratie avancée* » (*Le Messager*, n° 241, 1992 : 6).

⁴ La plus haute chaîne montagneuse du Cameroun : environ 4 100 m.

– *Démocratie de sous-développement* : semblant de démocratie initiée par l'occident pour l'Afrique. Ce type de démocratie encourage la « démocratisation » de l'Afrique tout en maintenant le continent sous le parapluie des intérêts occidentaux.

[21] « Il est temps que tous les leaders politiques comprennent que les occidentaux nous proposent une *démocratie de sous-développement*. C'est pourquoi, il est hasardeux de surestimer leurs incitations officielles à l'ouverture démocratique » (*Challenge Hebdo*, n° 44, 1991 : 11).

– *Démocratie du gari*⁵ : démocratie dont le principe fondamental est le pot-de-vin.

[22] « Ils vont être contents les Camerounais de savoir que dans l'opposition on ne pratique pas la *démocratie du gari* mais l'autre démocratie, la vraie » (*Le Messenger*, n° 230, 1991 : 11).

– *Démocratie du ventre* : (synonyme de ventrocratie).

[23] « Enfin pour vous prouver que nous sommes d'honnêtes citoyens, le peuple ira aux urnes pour entériner la *démocratie du ventre* » (*Challenge Hebdo*, n° 48, 1991 : 2).

– *Démocratie tropicale* : « démocratie » de type africain caractérisée par le fait que la minorité au pouvoir décide pour le peuple et au nom du peuple sans l'avis du peuple.

[24] « Cette intense activité inventrice n'a pas empêché Paul Biya [...] de se précipiter à l'Élysée pour se faire remettre son certificat de meilleur élève en *démocratie tropicale* » (*Le Messenger*, n° 227, 1991 : 2).

– *Conférence nationale* : cette expression est de toute évidence la plus employée au Cameroun de 1990 à 1992.

1- Selon l'opposition politique, il s'agit d'un cadre institutionnel de débat franc sur la vie sociopolitique camerounaise, sur les différends qui animent les Camerounais, afin de jeter les bases d'une véritable démocratie.

2- Selon le pouvoir, un cadre de débat existe déjà, représenté par l'Assemblée nationale. C'est pourquoi « la *conférence nationale* est sans objet », car elle est inconstitutionnelle. (*dixit* Paul Biya, chef de l'État camerounais, le vendredi 22 mars 1991).

[25] « La *conférence nationale* se présente comme un impératif incontournable aujourd'hui. C'est le seul forum où seront évalués sans complaisance tous les échecs du passé et s'élaboreront les bases solides d'une société véritablement démocratique » (*Le Messenger*, n° 239, 1991 : 13) ;

[26] « Oser évoquer une quelconque constitutionnalité de la *conférence nationale* relève d'une paresse et d'une malhonnêteté intellectuelle » (*Le Messenger*, n° 233, 1991 : 17) ;

[27] « Je crois que la question n'est plus de savoir si nous devons tenir une *conférence nationale* mais plutôt de savoir ce que doivent être les thèmes de référence, les personnes et la composition des délégations d'une telle conférence » (*dixit* Salomon Tandeng

⁵ le gari est un plat camerounais (généralement bon marché) fait à base de tubercules de manioc.

Muna, ancien président de l'Assemblée nationale camerounaise, le 6 mai 1991 ; in *Le Messenger*, n° 228, 1991 : 4).

– *Vandales* : leaders politiques de l'opposition qui encouragent la protestation et la désobéissance civile.

[28] « Aux groupuscules de *vandales* qui réclamaient à cor et à cri l'amnistie générale et inconditionnelle, le chef de l'État avait posé une condition [...] » (*Le Messenger*, n° 226, 1991 : 2).

– *Opposition des encaisseurs* : chefs des partis de l'opposition qui acceptent le compromis avec le régime en place moyennant généralement un poste de responsabilité dans les hautes sphères de l'État.

[29] « Le président-fondateur du Parti de la solidarité du Peuple (P.S.P) [...] préfère, pour sa survie personnelle et probablement celle du Cameroun, *l'opposition des encaisseurs* » (*Le Messenger*, n° 239, 1991 : 2).

– *Casseurs* : militants des partis d'oppositions, manipulés et spécialisés dans la destruction des biens du peuple camerounais.

[30] « Un infirme du micro essayait d'expliquer que les *casseurs* devaient se poser la question de savoir qui aurait plus à perdre en cas de destruction » (*Challenge Hebdo*, n° 28, 1991 : 5).

– *Opposition alimentaire* : chefs de partis politiques qui ne représentent qu'eux-mêmes et n'attendent que leur part du « gâteau ».

[31] « Le chef de l'État a maintenant en face de lui des interlocuteurs endurcis, prêts à monter les enchères... et, tout autour, des caciques de *l'opposition alimentaire* pour freiner le processus » (*Challenge Hebdo*, n° 50, 1991 : 5).

– *Multipartisme à parti unique* : dictature du parti au pouvoir dans un contexte de démocratie pluraliste.

[32] « Votre *multipartisme à parti unique* a du charme. À le contempler en action, on n'a vraiment pas le temps de s'ennuyer » (*Le Messenger*, n° 229, 1991 : 2).

En effet, la description analytique des néologies susmentionnées est révélatrice non seulement des faits de créativité du discours politique au Cameroun de 1990 à 1992, mais rend compte aussi de l'environnement psychologique et sociopolitique conflictuel d'alors. Ainsi, peut-on constater que l'échange langagier n'était pas une activité pacifique, mais plutôt un espace de tension et d'agression, et le changement langagier n'était que le reflet des antagonismes qui régissaient les relations interpersonnelles. Il ne saurait d'ailleurs en être autrement lorsque c'est dans « l'arène » politique que le langage fonctionne, « lieu fait lui-même d'antagonismes, traversé par de lourds enjeux de pouvoir et qui de plus est médiatisé à outrance dont objet de toutes les spectacularisations » (Boyer, 1991 : 82).

Quoi qu'il en soit, à travers les créations néologiques, on note de la part des acteurs politiques camerounais d'abord comme une politique-spectacle, avec une volonté de séduction et de manipulation du peuple. Ces derniers cherchent à influencer par des moyens qui

s'apparentent beaucoup plus au coup de force langagier qu'à la persuasion. En plus, ces mots constamment répétés sont demeurés dans la mémoire des Camerounais comme des paradigmes dignes du bourrage de crâne, et la manipulation de l'opinion publique est devenue une préoccupation centrale des acteurs politiques ; ce qui favorise une véritable guerre psychologique, surtout que chaque partie (le régime en place et l'opposition) est consciente que si l'opinion publique est avec elle, rien ne peut lui résister, et que sans l'opinion, tout pouvoir est impuissant.

Enfin, chaque camp tient un discours de type autoritaire et fait de l'autre le responsable des malheurs du peuple camerounais. À cet effet, selon le bord politique, on construit « une image diabolisante de l'autre » (Windich 1987 : 39), pour ainsi avoir un rayonnement sur l'espace sociopolitique national voire international ; d'où la nécessité pour les hommes politiques de venir, parés de leurs plus beaux atours, s'expliquer, se justifier, « se mettre à nu »... se vendre dans le marché médiatique où le lecteur camerounais ou étranger vient faire « un shopping démocratique » (Ghiglione 1991 : 54).

3. CONCLUSION

Dans cette communication, nous avons présenté et analysé, sur la base de la créativité néologique, un ensemble de lexies et expressions relevant du vocabulaire politique au Cameroun de 1990 à 1992. Ces néologies se chargent d'une bonne dose de connotations expressives qui révèlent les oppositions politiques au Cameroun au cours de la période étudiée. À cet effet, ces mots sont à l'image de la fracture sociopolitique qui, elle-même, renvoie aux antagonismes idéologiques. Pour accéder donc à la complétude sémantique de ces mots, il faut nécessairement intégrer le contexte extralinguistique par opposition au contexte linguistique. Ce contexte extralinguistique est spécifique et s'enracine dans un espace, dans un temps, avec une vision du monde particulière. En ce qui nous concerne, le contexte extralinguistique est incontestablement un élément pertinent pour le décodage et par ricochet pour une réinterprétation efficiente du signe linguistique.

BIBLIOGRAPHIE

- Boyer (H.), 1991 : *Éléments de sociolinguistique*, Paris, Dunod.
- Cusin-Berche (F.), 1999 : « Lexique en mouvement : création lexicale et production scientifique », dans *Langages*, n° 136, pp. 5-26.
- Dubois (J.), 1962 : *Le Vocabulaire politique et social en France de 1869 à 1872*, Paris, Larousse.
- Dubois (J.) (et al.), 1994 : *Dictionnaire de linguistique*, Paris, Larousse.
- Fosso, 1999 : « Créativité lexicale sur le campus universitaire de Yaoundé I : étude du champ lexical de la sexualité », dans *Le Français en Afrique noire*, revue du réseau des Observatoires du Français contemporain en Afrique, n° 13, (A. Queffelec, dir.), Paris, Didier Érudition, pp. 47-57.
- Fuch (Cl.), 1984 : « Le Sujet de la théorie énonciative d'Antoine Culioli : quelques repères », *DRLAV*, n° 30, Paris pp. 45-53.
- Ghiglione (R.), 1991 : *Je vous ai compris ou l'analyse des discours politiques*, Paris, A. Colin.
- Gumperz (J.), 1979 : *Sociolinguistique interactionnelle*, Paris, l'Harmattan.

Heller (M.), 2002 : *Éléments d'une sociolinguistique critique*, Paris, Didier.

Lehmann, (A.) et Martin-Berthet, (F.), 2003 : *Introduction à la lexicologie. Sémantique et Morphologie*, Paris, A. Colin.

Martinet (A.), 2003 : *Éléments de linguistique générale*, Paris, A. Colin.

Nga Ndongo, (V.), 1989 : « Le journal camerounais comme espace sociologique », dans *Annales de la F.L.S.H*, Université de Yaoundé, pp. 3-36.

Salminen (A. N.), 2000 : *La Lexicologie*, Paris, A. Colin.

Windsich (U.), 1987 : *Le K.O. verbal. La Communication conflictuelle*, Paris, l'Âge d'Homme.

LE DICTIONNAIRE DES COLLOCATIONS EN LIGNE

Antonio González Rodríguez

Département de traduction, interprétation et langues modernes

Universidad Europea de Madrid

Madrid, Espagne

Résumé : *les dictionnaires reprenant des structures binaires existent depuis fort longtemps et dans différentes langues. Le dictionnaire des collocations en ligne, disponible sur l'Internet, repose sur six combinaisons récurrentes de la langue française, puisées dans la presse, dans des romans et des essais. Cet outil est destiné aux traducteurs, aux interprètes, aux apprenants et à toute personne qui doit produire un texte en français.*

Mots-clés : *dictionnaire, collocations, traduction, français langue étrangère (FLE)*

1. INTRODUCTION

Le dictionnaire des collocations en ligne suit les brisées d'ouvrages similaires publiés dans différentes langues et pays. Un petit tour d'horizon de ces productions permettra de dégager les similitudes et les différences afin de mieux souligner l'originalité de ce dictionnaire des collocations en ligne.

2. DICTIONNAIRES DE COLLOCATIONS MONOLINGUES

En 1759, le Révérend Père Daire publie à Lyon, chez Pierre Bruyset Ponthus, *Les Épithètes françoises* rangées sous leurs substantifs, ouvrage utile aux Poètes, aux Orateurs, aux jeunes gens qui entrent dans la carrière des Sciences, & à tous ceux qui veulent écrire correctement, tant en Vers qu'en Prose. L'auteur propose une liste de substantifs accompagnés des adjectifs qui les qualifient. L'ouvrage n'aborde qu'une seule combinaison, à savoir « substantif + adjectif ». On retiendra que l'auteur a prévu un système de renvoi qui permet au lecteur de voyager utilement dans le dictionnaire.

En 1999, Georges-Charles et Mauricette Lecomte, père et fille, publient à compte d'auteur le *Dictionnaire des épithètes françaises*. À l'usage de tous ceux qui écrivent et qui aiment les mots. D'un point de vue formel, les auteurs ont classé les substantifs par ordre alphabétique tout comme les épithètes qui les accompagnent. Il est à signaler qu'ils ont également ajouté quelques segments récurrents. Ainsi, sous l'entrée « immeuble », on pourra lire, outre les adjectifs, des segments tels qu'« immeuble de belle apparence », « immeuble mis sous séquestre », « immeuble mis en vente » et « immeuble de rapport ». Formellement, les auteurs ne font pas la différence et classent donc ces segments avec les adjectifs. Un système de renvoi est également prévu, même s'il n'est pas systématique.

Jacques Beauchesne publie, en 2001, le *Dictionnaire des cooccurrences*, riche de quelque 4 200 entrées ou substantifs. L'auteur propose systématiquement trois types de combinai-

sons : « substantif + adjectif », « verbe ou locution verbale + substantif » et « substantif + verbe ». Ces catégories sont lâches, car il est difficile, et c'est là le problème majeur des collocations, de classer ces dernières dans des structures intangibles. Ainsi, dans la combinaison « substantif + adjectif », on pourra lire, sous l'entrée « imagination », les segments suivants : « en délire », « (jamais) prise en défaut », « qui fonctionne bien/rapidement » « sans cesse renouvelée », « sans bornes/frein/limites ». Cet ouvrage est, pour l'heure, le dictionnaire le plus complet en langue française.

Javier Boneu publie, en 2001, *El diccionario euléxico para expresarse con estilo y rigor*. Les substantifs espagnols, qui tiennent lieu de base ou de mot-clé, sont classés par ordre alphabétique. L'auteur propose trois combinaisons : « substantif + adjectif », « verbe + substantif » et « substantif + verbe ». Un système de renvoi est également prévu.

En 2004, Ignacio Bosque, professeur à l'université Complutense de Madrid, publie *Redes, diccionario combinatorio del español contemporáneo*. Cette somme, qui compte 1 839 pages et une introduction de 174 pages, est sans conteste l'ouvrage le plus complet et le plus complexe. Tous les ouvrages cités ci-dessus ne présentaient que des substantifs en guise d'entrée. L'auteur a prévu cinq types d'entrées, le substantif, le verbe, l'adverbe, l'adjectif et des locutions, ce qui donne quelque 10 combinaisons : « substantif + adjectif », « substantif + locution adverbale », « adjectif + substantif », « verbe + adverbe », « verbe + substantif », « verbe + locution adverbale », « adverbe + adjectif », « adverbe + verbe », « locution prépositive + substantif », « locution adverbale + verbe ». Cet ouvrage est doté d'un système de renvoi et de nombreux exemples puisés dans la presse espagnole et hispano-américaine.

En 1997, Morton Benson, Evelyn Benson et Robert Ilson publient *The BBI Dictionary of English Word Combinations*. Cette édition remaniée, riche de 18 000 entrées et de 90 000 collocations, propose trois types d'entrées ou bases : le substantif, l'adjectif et le verbe. Les auteurs considèrent 9 combinaisons : « verbe + substantif », « verbe + préposition », « substantif + adjectif », « substantif + préposition », « substantif + substantif », « substantif + verbe », « locution prépositive + substantif », « adjectif + préposition », et des expressions idiomatiques. Les auteurs n'ont pas prévu de système de renvoi.

3. DICTIONNAIRES DE COLLOCATIONS BILINGUES

Outre les dictionnaires monolingues, on trouvera aussi dans le commerce des dictionnaires de collocations bilingues. C'est le cas du *Dictionnaire des collocations chinois-français*, publié en 2002. Les substantifs, qui tiennent lieu d'entrée, sont classés par ordre alphabétique et alimentent les combinaisons suivantes : « locution verbale + substantif », « verbe + substantif », « substantif + adjectif » et les traductions correspondantes en chinois. Il n'y a pas de système de renvoi.

En 1979, Boris Vassilievitch Bratous publie le *Slovosotchetania rousskovo iasika*, un dictionnaire reprenant les collocations les plus usuelles en russe et leurs traductions en français. Chaque entrée, toujours un substantif, présente une structure ternaire : « substantif

+ adjectif », « substantif + substantif » et « verbe + substantif ». Il n'y a pas de système de renvoi.

4. MANUELS D'APPRENTISSAGE

Qui veut apprendre une langue étrangère doit en maîtriser les collocations et les apprendre par cœur. En 1970, Gak et Lvin publient le *Kyrs perevoda – fransuskie iasik*. Ce manuel, destiné aux apprentis traducteurs russes, se compose de 53 leçons qui obéissent toutes à une même structure : un texte en français ou en russe sur un sujet bien précis, par exemple des négociations commerciales entre deux États. Des phrases-types en français et en russe portant sur lesdites négociations. Un commentaire grammatical sur le texte et sur les structures utilisées dans les deux langues. Et toute une batterie d'exercices qui tournent autour du sujet étudié. En fin de leçon, les auteurs proposent un petit glossaire. Il ne s'agit pas d'une liste de mots, mais de combinaisons ayant pour base un mot-clé de la leçon. Les auteurs proposent 6 structures : « substantif + substantif », « substantif + adjectif », « verbe + substantif », « locution verbale + substantif », « substantif + verbe » et « locution prépositive + substantif ».

En 2003, Gak et Grigoriev publient *Teoria y praktika perevoda, fransuskie iasik*. Bien que ce manuel présente beaucoup de points communs avec l'édition de 1970, le dictionnaire en fin de leçon ne repose plus sur un seul mot-clé, mais sur plusieurs mots étudiés dans la leçon. Pour autant, les mots sont souvent présentés avec leurs collocatifs.

Bouscaren et Lab publient, en 1998, *Les Mots entre eux, Words and their collocations*. Ce vocabulaire regroupe par thèmes (le corps, l'enseignement, la santé, les courses, etc.) environ 30 000 mots et leurs collocatifs. Dans chaque chapitre, les mots anglais sont suivis de la traduction française et des collocations anglaises. Afin de bien les fixer, les auteurs ont prévu des exercices d'application.

5. LE DICTIONNAIRE DES COLLOCATIONS EN LIGNE

C'est en 1990 que nous commençons à classer, dans des fichiers informatiques, les collocations que nous rencontrons principalement dans *Le Monde*, *Le Monde diplomatique*, *Courrier international*, *Science et Vie*, *Le Canard enchaîné*, *Lire*, *Le Magazine littéraire* et, plus généralement, dans des romans et des essais. Au fil de ces années de travail empirique, nous avons établi plusieurs catégories qui sont celles que nous venons de voir et que nous allons présenter par le menu.

4.1. Verbe + nom

Dans un premier temps, par souci stylistique, nous avons cherché les verbes qui accompagnent systématiquement les noms et qui remplacent les verbes dits « vides » :

- | | | |
|-----------------------|---|----------------------------|
| [1] faire une annonce | → | passer une annonce |
| [2] faire un appel | → | lancer un appel |
| [3] faire ses bagages | → | boucler, plier ses bagages |

- [4] faire un catalogue → dresser un catalogue
 [5] faire un métier → exercer, pratiquer un métier

Dans certains cas, il ne s'agit plus de remplacer les verbes vides ou verbes supports, mais plus simplement de verbaliser le nom. N'oublions pas que le traducteur est, à l'occasion, tenu de jouer, à la hausse ou à la baisse, avec le nombre de mots pour des raisons d'espace :

- [6] apporter une aide revient à dire aider
 [7] pousser un aboiement revient à dire aboyer
 [8] formuler une accusation revient à dire accuser
 [9] accorder une amnistie revient à dire amnistier
 [10] donner son approbation revient à dire approuver

Plus généralement, nous avons repris tous les verbes qui sont susceptibles, en compagnie du nom, d'apporter un sens singulier à la collocation :

- [11] contrat conclure, décrocher, dénoncer, empocher, emporter, entériner, exécuter, honorer, négocier, passer, rédiger, remplir, renégocier, résilier, rompre, signer.

Outre la combinaison « verbe transitif direct + nom », nous travaillons également la structure « nom + verbe intransitif » :

- [12] une flamme vacille
 [13] un fleuve gronde
 [14] une explosion retentit
 [15] l'Histoire bégaye

4.2. Nom + adjectif

La combinaison « nom + adjectif » est probablement l'une des plus productives. On retiendra que les dictionnaires généraux ne peuvent sanctionner tous ces usages qui relèvent bien souvent de la création subjective. Comme le dictionnaire des collocations est, avant toute chose, un outil pour produire, il se doit de reprendre l'ensemble des adjectifs qui accompagnent naturellement les substantifs. Le substantif « précision » est souvent suivi des adjectifs suivants :

- [16] chirurgicale, diabolique, maniaque, mathématique, etc.

4.3. Les segments récurrents

Il existe, au-delà des structures binaires que nous venons de voir, des segments récurrents qui répondent à différents noms selon les spécialistes en la matière. Ces segments enchâssent une base, qui est l'élément transparent, d'une langue à l'autre ou dans sa propre langue, qui nous sert de point de référence pour le classement dans nos fichiers ou dans notre dictionnaire sur la Toile. Mais prenons le cas de « justice » pour éclairer notre propos.

Le traducteur qui doit traduire le mot ou le concept espagnol « justicia » pourra s'arranger des trois structures binaires décrites ci-dessous :

verbe + nom :

[17] justice bafouer, fuir, obtenir, réclamer, rendre, saisir

nom + verbe :

[18] justice trancher

nom + adjectif :

[19] justice clémente, expéditive, réparatrice, sociale, sommaire, spectacle, tardive

Or, bien souvent, le texte exigera l'utilisation de structures plus complexes :

[20] l'exercice de la justice

[29] être inquieté par la justice

[21] se frotter à la justice

[30] porter une affaire en justice

[22] menacer d'un recours en justice
justice

[31] avoir maille à partir avec la

[23] échapper à la justice

[32] une justice à sens unique

[24] la justice a la main lourde

[33] la justice le rattrapera

[25] l'acharnement de la justice

[34] des démêlés avec la justice

[26] une obstruction à la justice

[35] comparaître devant la justice

[27] être traduit en justice
tice

[36] être en délicatesse avec la jus-

[28] être blanchi par la justice

[37] être épris de justice

4.4. Verbe + complément

De même que certains noms appellent systématiquement un adjectif, de même les verbes, certains plus que d'autres, sont suivis d'un adverbe ou d'un complément. Une fois de plus, tout est question de fréquence et, partant, d'usage. Ainsi le verbe « critiquer » se construit-il souvent avec :

[38] critiquer à mots couverts, âprement, durement, machinalement, ouvertement, sèchement, vertement, vivement.

4.5. Adjectif + nom

Le traducteur doit à l'occasion se raccrocher au terme le plus transparent, une manière de planche de salut. Or, dans certains cas, c'est l'adjectif qui l'aidera à rendre correctement le nom. Il nous souvient du titre d'un article paru dans le *New York Times* début 2002 : « Nuclear Posture Review ». La consultation d'un dictionnaire bilingue ne sera pas d'un grand secours en pareil cas.

Et pour cause, ces dictionnaires sont utiles dans l'apprentissage des langues. Or le « traducteur est supposé les connaître avant que de traduire », selon les termes de Valéry Larbaud (1973 : 88). L'adjectif français « nucléaire » est souvent précédé du nom « doctrine ». En d'autres termes, « nucléaire » appelle « doctrine », et l'on parle bien d'une « doctrine nucléaire ». Nous avons donc repris ces adjectifs et les noms qui les accompagnent systématiquement et qui sont moins transparents, à l'heure de traduire ou de rédiger, que l'adjectif lui-même.

Les adjectifs espagnols *presupuestario*, *mediático*, *tecnológico* ne posent aucun problème de traduction : « budgétaire », « médiatique » et « technologique ». En revanche, les noms qui les précèdent peuvent donner du fil à retordre à moins que l'on ne connaisse ce qui suit :

[39] budgétaire	arbitrage, assainissement, austérité, choix, contrainte, coupe, dérapage, discipline, rallonge, restriction, rigueur, trou ;
[40] médiatique	acharnement, battage, conditionnement, couverture, déballage, emballage, relais, retentissement, sirène, tapage, tintamarre ;
[41] technologique	avancée, bijou, bond, fossé, gouffre, innovation, pari, percée, prouesse, raffinement, retombée, saut, tour de force.

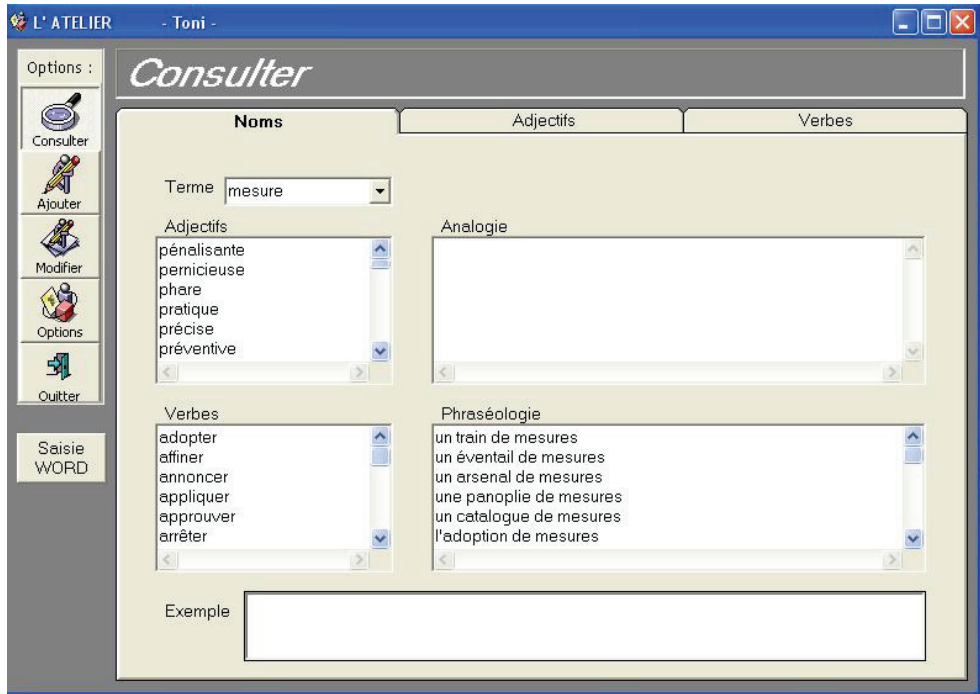
Nous avons donc établi 6 catégories ou combinaisons qui intéresseront tous ceux qui doivent rédiger un texte en français :

1. verbe + nom
2. nom + verbe
3. nom + adjectif
4. segments récurrents
5. verbe + adverbe ou complément
6. adjectif + nom

4.6 L'Atelier, une base de données de collocations


Ces collocations ont été enregistrées, en 2000, dans une base de données *Access* de Microsoft® installée sur le disque dur de notre ordinateur. À rebours des différents fichiers *Word* où nous consignons patiemment les collocations que nous glanons de-ci de-là, l'utilisation de cette base de données *Access* permet, à la faveur d'une interface utilisateur réalisée avec *Visual Basic* de Microsoft®, un accès direct et immédiat aux collocatifs. Cette base de données est en contact direct avec le logiciel *Word*, puisqu'il nous suffit de cliquer sur un mot d'un texte pour ouvrir la base directement et immédiatement à la bonne fiche.

L'Atelier, nom que nous lui avons attribué, car la langue est notre outil de travail, présente trois onglets correspondant aux trois bases (substantif, adjectif et verbe) sur lesquelles reposent les six combinaisons que nous travaillons. Conscient du fait que l'on ne traduit pas que des mots, nous avons ajouté une fenêtre « analogies » qui propose des chemins de traverse sur la route du sens. Cette base de données, alimentée régulièrement, tout comme les fichiers *Word* et la version en ligne, est la roue de secours qui nous dépanne lorsque la connexion à la Toile fait défaut.



4.7. Le dictionnaire des collocations en ligne

Enfin, le 14 février 2004, nous avons publié sur l'Internet (www.tonitraduction.net) toutes les collocations sous la forme d'un dictionnaire des collocations, qui compte quelque 27 400 combinaisons. Le site Web présente trois hyperliens, soit les trois bases (substantif, adjectif et verbe) qui servent de porte d'entrée à toute recherche. Il est, dans certains cas, possible de cliquer sur les collocatifs ; moyennant les hyperliens, l'utilisateur pourra sauter d'une collocation à l'autre au gré de ses besoins. Le site Web propose aussi des exercices destinés à toute personne soucieuse de maîtriser les collocations de la langue française pour diverses raisons. Que l'on songe à l'apprenant étranger qui pourra, dès lors qu'il les maîtrise, bâtir une phrase, au traducteur, toujours en quête d'un surcroît de productivité, à l'interprète qui, par le truchement des collocations, pourra anticiper le sens et au producteur de textes en panne d'imagination.

Direction <http://www.tonitraduction.net>  fr

DICTIONNAIRE DES COLLOCATIONS

[Accueil](#) [Mode d'emploi](#) [Contact](#) [Exercices](#)

[Adjectifs](#) [Substantifs](#) [Verbes](#)

 [Afficher tout](#)

Page sur 1

Enregistrements 1 à 1 sur 1

Substantifs (*)

mesure

©Antonio Gonzalez Rodriguez, 2004

SUBSTANTIFS: mesure

Adjectifs	Verbes	Segments récurrents	Analogies
aberrante	adopter	adoption (l') de mesures	
appropriée	affiner	annonce (l') de mesures	
avantageuse	annoncer	applicabilité (l') d'une mesure	
coercitive	appliquer	arsenal (un) de mesures	
concrète	approuver	catalogue (un) de mesures	
contraignante	arrêter	durcissement (le) des mesures	
controversée	assouplir	éventail (un) de mesures	
corrective	battre	impact (l') des mesures	
cosmétique	concocter	mesure (une) d'accompagnement	
disciplinaire	donner	mesure (une) de bon sens	
discriminatoire	égrener	mesure (une) mal ficelée	
discutable	entériner	mesure (une) qui est étendue à	
dissuasive	énumérer	mesure d'austérité	
douloureuse	envisager	mesures (des) attendues de longue date	
draconienne	fausser	mesures (des) au cas par cas	
efficace	financer	mesures (des) d'urgence	
énergique	imposer (s')	mesures (des) de circonstance	
forte	préconiser	mesures (des) de coercition	
gadget	prendre	mesures (des) de pure forme	
immorale	prononcer	mesures (des) de régression sociale	
impopulaire	réclamer	mesures (des) de représailles	
inacceptable	retenir	mesures (des) de rétorsion	
incitative	saluer	mesures (des) en trompe-l'œil	
inefficace		mesures (des) restées lettre morte	

Ce dictionnaire présente trois avantages. D'abord, il est gratuit. Ensuite, il est évolutif en ce sens qu'il est alimenté régulièrement et épouse les tendances de la langue française. À l'inverse de l'édition papier, l'édition numérique est immédiate. De surcroît, à l'instar de tout programme informatique, il nous est loisible d'ajouter à tout moment une collocation, de la modifier ou de la supprimer, voire de remanier de fond en comble la présentation du dictionnaire, comme nous avons prévu de le faire afin de tirer un meilleur parti des potentialités de la base de données. Enfin, s'agissant d'un outil pour produire, sa présence sur le Net permet d'accroître la productivité du travail du producteur.

4.8. Conclusions

Le dictionnaire des collocations en ligne repose sur l'observation des combinaisons récurrentes de la langue française, à l'instar de ces dictionnaires combinatoires publiés en d'autres langues. Cette observation empirique nous a conduit à dégager six combinaisons. Cet outil, qui est destiné au traducteur, à l'interprète, à l'apprenant ou à tout producteur de textes, n'est pas un dictionnaire de compréhension de la langue française, mais bien un outil de production.

BIBLIOGRAPHIE

- Beauchesne (J.), 2001 : *Dictionnaire des cooccurrences*, Montréal, Guérin.
- Benson (M.), Benson (E.) et Ilson (R.), 1997 : *The BBI Dictionary of English Word Combinations*, Amsterdam/Philadelphie, John Benjamins Publishing Company.
- Boneu (J.), 2000 : *Diccionario euléxico para expresarse con estilo y rigor*, Barcelona, Editorial Juventud.
- Bosque (I.), 2004 : *Redes, diccionario combinatorio del español contemporáneo*, Madrid, Ediciones SM.
- Bouscaren (C.), Lab (F.), 1998 : *Les Mots entre eux. Word and their collocations*. Paris, Ophrys.
- Bratous (B. V.), 1979 : *Slovosochetania russkovo iasika*, Moscou, Russkie Iasik.
- Daire (R.P.), 1759 : *Les épithètes françaises*, Lyon, Pierre Bruyset Ponthus.
- Gak (V.), Lvine (Y.), 1970 : *Kurs perevoda, frantsusskii iasik*, Moscou, Mehdunarodnie Otnochenia.
- Gak (V.), Grigoriev (B.), 2003 : *Teoria y praktika perevoda, fransuskie iasik*, Moscou, Interdialect.
- González Rodríguez (A.), 2004 : *Dictionnaire des collocations*, <http://www.tonitraduction.net>.
- Larbaud (V.), 1973 : *Sous l'invocation de saint Jérôme*, Paris, Gallimard.
- Lecomte (G.-C.), Lecomte (M.), 1999 : *Dictionnaire des épithètes françaises*, Paris, Éditions françaises Cd-Romans.

TRADUIRE LES TERMES DE COULEUR : LA CHROMONYMIE EN BOTANIQUE À LA RENAISSANCE

Philippe Selosse

Université Lumière-Lyon II
CELL/UMR 5037 Institut d'histoire de la pensée classique
Lyon, France
selosse.philippe@wanadoo.fr

Résumé : *Le but de cette communication est méthodologique et épistémologique : montrer que pour comprendre et traduire des termes qu'occultent de nombreuses difficultés (langue de spécialité, langue morte, épistémè différente), il est nécessaire de les resituer dans leur paradigme de savoir par une mise en contexte multiple (référentielle, taxinomique, cognitive, épistémique et sociohistorique). L'exemple choisi est celui des termes de couleur (chromonymes) dans la botanique de la Renaissance.*

Mots-clés : *chromonymes, épistémè, taxinomie, prototypie, référentialisme*

1. INTRODUCTION

Rien de plus simple, semble-t-il, que la traduction des termes de couleur (désormais « chromonymes »), tant les référents semblent communs à toutes les cultures. Cette communication montrera pourtant que pour comprendre et traduire de tels termes, il est nécessaire de les resituer dans leur *épistémè* (Foucault 1966) par de multiples mises en contexte. L'exemple choisi pour la démonstration est celui des chromonymes dans la botanique de la Renaissance.

À la différence du Moyen-Âge qui se limitait pour l'essentiel à une compilation des livres des Anciens, la Renaissance se caractérise, dans le domaine des plantes, par un retour à l'observation de terrain. Il en résulte que les hommes de la Renaissance se trouvent en présence d'une multiplicité de plantes jusque-là inconnue, mais aussi en présence d'une multiplicité de couleurs nouvelles auxquelles les Anciens, comme Pline, n'avaient guère donné d'importance, car leurs notations descriptives portaient surtout sur la taille des plantes ou la forme des racines et des feuilles.

1.1. La signification des couleurs dans l'épistémè de la Renaissance

Cette multiplicité de couleurs portée par les plantes s'inscrit dans un paradigme particulier, à un moment où la conception des couleurs s'est fortement modifiée et enrichie par rapport à l'Antiquité et au Moyen-Âge, le triptyque de base « blanc, rouge, noir » s'étant étoffé de bien d'autres couleurs (Pastoureau 2000). C'est pourquoi les couleurs constatées dans les plantes sont porteuses de plusieurs dimensions dans l'épistémè de la Renaissance, la première d'entre elles étant religieuse : à l'instar de la diversité des plantes, la variété des couleurs est perçue comme une manifestation de la puissance divine, qu'on doit à ce titre révè-

rer. Autrement dit, les couleurs apparaissent comme une signature du divin dans la plante, ce qu'exprime par exemple le botaniste Caspar Bauhin à propos des tulipes :

[1] « En résumé, il existe une telle variété de couleurs parmi les Tulipes qu'encore aujourd'hui, chaque année, il en naît dont on n'a jamais vu de semblables : et c'est ainsi que DIEU manifeste dans ces fleurs mêmes son admirable puissance » (Bauhin 1623 : 63).

La deuxième dimension des couleurs est anthropocentrique. L'homme est posé au centre de la création et les plantes sont conçues comme lui étant « utiles », d'un point de vue officinal, culinaire, fourrager ou même esthétique : reprenant un argument de Pline, de nombreux botanistes rappellent que les plantes sont agréables à l'homme, particulièrement par leurs couleurs qui « réjouissent l'œil de l'homme », ce qui revient à dire que la couleur établit une relation entre le microcosme humain et le macrocosme des plantes. Enfin, une troisième dimension, ontologique, caractérise les couleurs : la couleur est alors perçue comme la « note caractéristique » de chaque plante, ce qui lui est propre et la définit en soi et par rapport aux autres au sein de la création. L'appréhension des couleurs se situe donc dans un mode descriptif qui relève d'un paradigme religieux ou métaphysique et qui diffère entièrement de celui pratiqué par la botanique moderne qui, elle, est descriptive à des fins taxinomiques, parce qu'inscrite dans un paradigme de connaissance.

1.2. L'interprétation des couleurs : un problème épistémologique

Si la signification globale des couleurs des plantes dans l'épistémè de la Renaissance est acquise, il n'en va pas de même de l'interprétation et de la traduction de chacun des chromonymes, qui soulèvent de nombreuses difficultés épistémologiques qu'on peut énumérer ainsi :

- traduction d'une langue morte : tout, ou presque, ce qui concerne les plantes à la Renaissance est écrit en néolatin ;
- double traduction : traduire en français les chromonymes néolatins, c'est opérer une deuxième traduction, après la première réalisée par les botanistes eux-mêmes, de leur vernaculaire (français, allemand, italien...) en néolatin ;
- distance diachronique : la représentation des couleurs a fortement évolué entre la Renaissance et le XXI^e ;
- hétérogénéité conceptuelle : les couleurs ne sont pas perçues objectivement mais socialement, culturellement (Pastoureau 2000), de sorte qu'il y a autant de couleurs que de cultures propres aux botanistes européens ;
- hétérogénéité taxinomique : les couleurs attribuées à une plante ne visent pas toujours la même espèce selon les auteurs, si bien qu'une plante peut être caractérisée par de multiples chromonymes ;
- absence de couleurs : le XXI^e s. n'a aucun référent à partir duquel avoir une idée de la couleur dénotée. En effet, les plantes dont nous disposons ne se trouvent que dans des herbiers secs, dans lesquels la couleur est passée, ou ne sont figurées que dans des gravures en noir et blanc.

Deux questions se posent alors : comment comprendre les couleurs exprimées en néolatin, malgré tant de difficultés ? Comment en rendre compte dans notre vernaculaire français moderne ? Car il est nécessaire de comprendre et de rendre compte de ces couleurs, puisqu'elles sont essentielles dans la description de la plante et dans le paradigme religieux de la Renaissance. Pour les terminologues, la solution réside dans le recours à la plante-référent, lorsque celle-ci peut être identifiée en taxinomie moderne, ce qui permet une identification des couleurs signifiées. Ce que je voudrais montrer ici, c'est combien cette tentation réaliste du référentialisme risque de fausser l'analyse des faits linguistiques, en l'occurrence des chromonymes.

2. LE « RÉEL », UNE MISE EN CONTEXTE POSITIVISTE

Malgré les difficultés signalées *supra*, il est parfois possible d'identifier une plante et de voir « directement » ses couleurs. Le cas, rare, se présente pour des plantes que l'on trouve dans des éditions coloriées à la main à la Renaissance, souvent d'après nature (le coloriste intervient sur un ouvrage déjà imprimé mais travaille à partir de la plante-référent qui a servi à la gravure et qu'il a sous les yeux), mais plus souvent encore d'après l'interprétation du chromonyme que se fait le coloriste français, allemand, italien... qui ne dispose d'aucun échantillon (et encore moins de la plante originelle) lui permettant de connaître la couleur exacte propre à la plante – autant dire, dans ce dernier cas, que la référence à la plante est une aide peu fiable. On partira cependant, pour notre démonstration, du cas de figure le plus fiable, avec une plante bien identifiée, portant une couleur stable et toujours représentée de la même façon : en l'occurrence, le lys orangé (*Lilium bulbiferum* L. en nomenclature moderne) que les éditions coloriées, soit d'après nature (Fuchs 2001), soit selon l'interprétation du coloriste (Maignan 1549 ; Anonyme 1549) représentent toutes en jaune-orangé. Or, ce Lys est qualifié au XVI^e. de multiples façons :

[2a] « *Lilium purpureum* » = « lys pourpre » chez Lobel et Daléchamp (français), Dodoens (néerlandais), Turner (anglais) et Besler (allemand) ;

[2b] « *Lilium rubrum* » = « lys rouge » chez Brunfels et Fuchs (allemands) et Gesner (suisse) ;

[2c] « *Lilium rufum* » = « lys roux » chez Bock (allemand) Dodoens (néerlandais) ;

[2d] « *Lilium croceum* » = « lys jaune safran » chez Lonicer (allemand), Caspar Bauhin (suisse) et Césalpin (italien) ;

[2e] « *Lilium aureum* » = « lys jaune d'or » chez Tabernaemontanus et Brunfels (allemands).

On se trouve en présence d'un chaos chromonymique qui ne permet pas à l'historien ou au botaniste moderne de déterminer la couleur de la plante : pourquoi de telles variations terminologiques, qui ne s'expliquent pas par une diversité culturelle, comme le montre le fait que les Allemands utilisent toutes ces dénominations ?

2.1. Le système des couleurs selon les ethnolinguistes

On peut chercher une première réponse à cette question en optant pour le référentialisme de certains ethnolinguistes, qui considèrent qu'il y a sept stades dans l'évolution des cultures et dans la perception et la dénomination des couleurs qui seraient « universellement perçues », depuis le blanc et le noir (premier stade), jusqu'à l'orange et au gris (septième stade, celui des civilisations censées être les plus développées) :

[3] graduation dans l'évolution des couleurs : (1) blanc et noir > (2) rouge > (3) vert/jaune > (4) bleu > (5) brun > (6) pourpre/rose > (7) orange/gris (Berlin et Kay 1969)

Selon cette théorie évaluative, la variabilité des notations chez les botanistes de la Renaissance pour caractériser le lys orangé traduirait une extrême confusion jointe à l'incapacité à nommer une couleur alors nouvelle (l'orange). Je réfuterai ici brièvement cet argument, très souvent rappelé à propos de la couleur orange. D'une part, lorsque les botanistes de la Renaissance veulent référer à une nouvelle couleur, ils utilisent des périphrases comparatives, comme le montrent les exemples [4]- [9] chez Caspar Bauhin (1623) :

[4] « *Tulipa dubia media sulphurei coloris.* »

« Tulipe douteuse, moyenne, de la couleur du soufre. »

[5] « *Linaria tenuifolia, aeruginei coloris.* »

« Linaire à feuilles minces, de la couleur de la "rouille de cuivre" ».

[6] « *Orchis papilionem expansum herbacei coloris referens.* »

« Orchis représentant un papillon aux ailes étalées de la couleur de l'herbe. »

[7] « *Clematitis Malabarensis foliis Vitis, colore dracunculi.* »

« Clématite de Malabar à feuilles de vigne, de la couleur de l'estragon. »

[8] « *Aspalathus colore buxi.* »

« Aspalathus de la couleur du buis. »

[9] « *Pisum exoticum ochrae colore.* »

« Pois exotique de la couleur de l'ocre. »

D'autre part, comme on le voit encore chez Caspar Bauhin (1623), ces périphrases comparatives valent aussi pour la couleur nouvelle orange [10] et sont même concurrencées par l'adjectif lexicalisé *orangeus* [11] :

[10] « *Paeonia pomi arantii colore.* »

« Pivoine de la couleur de la pomme orange. »

[11] « *Anemone latifolia flore orangei coloris.* »

« Anémone à larges feuilles à fleur de couleur orange. »

Si la couleur orange était visée pour ce que nous appelons « Lys orangé », les botanistes du XVI^e. auraient eu recours aux périphrases comparatives. S'ils n'ont pas utilisé ce moyen qui était à leur disposition, c'est donc que leurs dénominations visaient une autre couleur ou visaient la couleur orange sous une autre dimension perceptive.

2.2. Le système des couleurs selon les physiciens

Une deuxième solution référentialiste se présente alors, sous l'aspect de la classification internationale des couleurs de Munsell, qui définit chaque couleur selon trois attributs : teinte, luminosité et saturation (intensité). On tâchera de mesurer ici la pertinence de ces trois attributs pour comprendre la chromonymie. Selon ce système, on remarque d'abord que tous les termes choisis pour qualifier le « lys orangé » seraient dits intenses et sombres en français moderne :

[12a] *purpureum* = pourpre, *i.e.* rouge violacé = intense et sombre ;

[12b] *rubrum* = rouge soutenu, sombre ≠ *rubellum* = rouge faible, clair ;

[12c] *aureum/croceum* = jaune d'or/jaune safran = jaune très soutenu et sombre ;

[12d] *rufum* = roux = rouge très intense et foncé.

On note ensuite qu'une nouvelle dénomination, propre à Caspar Bauhin, va être adoptée dans l'ensemble de la communauté botanique et supplanter celles vues en [2a]- [2e], y compris chez tous ceux qui proposaient les chromonymes rappelés en [12a]- [12d].

[13] « *Lilium purpurocroceum* » (Bauhin 1623)

Or, cet adjectif composé permet deux choses : d'une part, la notation d'une forte saturation et d'une faible luminosité, traits qui sont communs aux deux interfixes (*purpuro-* et *croceum*), comme rappelé en [12a] et [12c] ; d'autre part, l'expression de deux teintes, le rouge (grâce à *purpuro-*) et le jaune (grâce à *croceum*), si bien que les deux couleurs primaires présentes dans l'orange sont signifiées. Grâce à la physique des couleurs, tout semble donc s'expliquer : c'est bien la couleur orange qui était visée ; le chaos des chromonymes d'avant la dénomination de Bauhin traduisait une perception de cette couleur non pas sous l'angle de la teinte, mais seulement sous celui de la saturation et de la luminosité.

2.3. Bilan critique

Ces deux approches référentialistes échouent cependant à rendre compte de la perception des couleurs à la Renaissance. Les ethnolinguistes présupposent une même perception des couleurs sous le seul aspect de la teinte ; or, les botanistes du XVI^e. ne perçoivent pas, dans le cas qui nous occupe, la teinte orange (sinon l'adjectif *orangeus* se serait imposé) mais une autre couleur unique quoique complexe et dont rend compte l'adjectif composé *purpurocroceum*. Autrement dit, traduire en [13] cet adjectif composé par l'adjectif simple « orange » serait affaiblir la complexité de la perception, d'où à mon avis la nécessité d'une traduction en français par un autre adjectif composé, « pourpre-safrané ». Les physiciens, quant à eux, présupposent une même conceptualisation des couleurs selon les trois modes de la teinte, de la luminosité et de la saturation, si bien que pour eux il y a une parfaite transparence linguistique et une identité évidente entre *purpureum* et pourpre, *croceum* et

jaune safran, etc. Or, est-on bien sûr que les deux adjectifs latins dénotent des teintes rouge et jaune ? qu'ils marquent une forte saturation et une faible luminosité ? De la réponse à ces deux questions dépend la validité de l'explication vue en section II.2. Or, rien d'autre que l'idée universaliste et réaliste d'une couleur pourpre immuable (ou d'une couleur safran invariante) ne vient confirmer que *purpureum* est perçu sous l'angle du rouge et dénote une couleur intense et sombre. Quand on sait, pourtant, que *purpureum* est défini par les lexicographes de la Renaissance (Nicot 1606) comme la couleur représentative des violettes (!), il est permis de douter de la démarche des physiiciens. Le contexte référentiel ne permet donc pas de trancher : au mieux, il permet de comprendre qu'une couleur particulière est dénotée par deux signifiants en composition adjectivale et que l'adjectif composé marque une variété de couleurs perçue.

Revenons-en donc à la traduction composée que je propose. Les linguistes latinistes estiment que la traduction par « pourpre-safrané » est à écarter, en raison de la syntaxe régressive qui est celle des composés latins ; ils proposent donc « safrané-pourpre ». Cette traduction est-elle meilleure et plus correcte ? Pour en juger, il faut donc se tourner vers d'autres contextes que le contexte référentiel.

3. LA MISE EN CONTEXTE, UNE APPROCHE CONCEPTUALISTE

Sur le plan strictement linguistique, le *corpus* informatisé de plus de 30 000 dénominations néolatines que j'ai constitué n'apporte aucune aide pour trancher. Il faut donc croiser les mises en contexte, autres que référentielle et linguistique, pour affiner l'interprétation vaguement dégagée précédemment. À cette fin, il me paraît nécessaire de faire appel à des mises en contexte relatives, non pas au référent, mais aux pratiques humaines et à leurs modes de conceptualisation, selon une hypothèse que je tiens pour fondamentale :

[14] « Le référent n'est jamais l'objet, réel ou imaginaire, abstrait ou concret, mais le concept de cet objet dans une langue donnée. Le linguiste n'a jamais pour objet d'étude le monde réel, mais ses conceptualisations et la mise en mots de ces conceptualisations » (Marie Luce Honeste, communication personnelle).

3.1. Pratique de la classification

Une première mise en contexte susceptible de nous aider consiste à voir comment les chromonymes s'intègrent dans la classification botanique de l'époque. Si l'on se réfère à ce contexte taxinomique, on s'aperçoit qu'il existe alors deux genres, *Lilium purpureum* et *Lilium croceum*. Une plante telle que [13] est rangée dans le genre *Lilium purpureum* et non dans le genre *Lilium croceum* : traduire en français par une syntaxe progressive (« pourpre-safrané ») permet de respecter, dans la dénomination propre à la Renaissance, le statut de la couleur pourpre qui est, comme je l'ai démontré ailleurs (voir Selosse à paraître), de signifier l'appartenance au genre *Lilium purpureum*. Par ailleurs, la plante [13] est rangée à la fin du genre *Lilium purpureum* et juste avant le genre *Lilium croceum* qui le suit : traduire par une syntaxe progressive permet de respecter, dans la dénomination, le statut de la couleur safran qui est de signifier la transition de *Lilium purpureum* à *Lilium croceum*. C'est là une dimension essentielle dans la classification botanique de l'époque que ce marquage d'une transition continue entre les plantes, nommée « chaîne des êtres ». Il

est inutile d'insister pour montrer que la traduction des physiiciens par un simple « orange » briserait cette double fonction de la couleur inscrite dans *purpurocroceum*.

3.2. Pratiques cognitives et épistémiques

Pour continuer à montrer que les données linguistiques ne permettent pas de discriminer un signifié sans l'apport de contextes et pour montrer l'importance d'autres contextes que le réel et la classification, j'aurai à présent recours à la construction syntaxique la plus fréquemment utilisée par Caspar Bauhin (1623) dans sa nomenclature pour exprimer les couleurs. Cette construction présente trois variantes [15a]- [15c] parfois utilisées toutes trois par un même botaniste pour une même plante – elles seront donc considérées comme équivalentes :

[15a] Prép. 'ex'+ N_[couleur 1] + Part. prés. _[couleur 2]

[16] *Verbascum nigrum flore ex luteo purpurascente.*

« Bouillon noir à fleur de jaune fauve tirant sur le pourpre. »

[15b] Adj. _[couleur 1] + Part. prés. _[couleur 2]

[17] *Cytisus hirsutus flore luteo purpurascente.*

« Cytise hirsute à fleur jaune fauve tirant sur le pourpre. »

[15c] Adj. composé [Adj. _[couleur 1] + Adj. _[couleur 2]]

[18] *Glyzyrrhiza sylvestris floribus luteopallescentibus.*

« Réglisse sauvage à fleurs jaune fauve tirant sur le jaune pâle. »

Sur le plan linguistique, on observe des adjectifs de couleur de forme différente, signifiant des couleurs de statut différent : l'emploi d'un adjectif simple pour signifier la première couleur (parfois introduite comme originelle par *ex*) indique qu'elle est présentée comme parfaitement réalisée ; l'emploi d'un adjectif dérivé de verbe, à désinence aspectuelle inchoative (morphème *-sc-* : *purpurascens*, *pallescens*) pour signifier la seconde couleur, indique qu'elle est présentée comme imparfaitement réalisée.

Pour affiner la compréhension de cette construction, il faut toutefois aller au-delà des seules données linguistiques et recourir à une interprétation contextuelle multiple. Le contexte référentiel, d'abord, nous apprend qu'il s'agit de plantes à fleurs jaunes aux étamines pourpres [16], à fleurs jaunes maculées de brun-pourpre [17] et à fleurs jaune pâle [18] : la construction s'applique donc aussi bien lorsqu'il s'agit de signifier deux couleurs distinctes que deux couleurs mêlées ou une seule couleur. Le contexte référentiel permet ainsi de rejeter l'idée que cette construction syntaxique signifierait une couleur unique ou dominante : il ne s'agit nullement de nuances, comme l'envisagent généralement les traducteurs. Par ailleurs, ce contexte révèle aussi une présentation spatialisée des couleurs, *i.e.* un regard qui va d'un point à un autre, mais parfois inversement sur le plan référentiel, la première couleur pouvant être périphérique [16] ou centrale [19] :

[16] *Verbascum nigrum flore ex luteo purpurascente.*

« Bouillon noir à fleur jaune fauve sur le pourtour et au cœur tirant sur le pourpre. »

[19] *Adonis hortensis flore minore atrorubente.*

« Adonis des jardins à fleur plus petite à cœur noir et tirant sur le rouge sur le pourtour. »

Le contexte référentiel est d'un apport certain, en ce qu'il montre l'imprévisibilité des données linguistiques, mais il reste insuffisant pour expliquer la quadruple combinatoire de couleurs révélée en [16]- [19].

Un deuxième contexte, le contexte classificatoire, élucide un peu plus le statut de la construction syntaxique [15a]- [15c] : il nous montre que de telles plantes se situent en fin de genre, *i.e.* qu'elles appartiennent au genre caractérisé par la première couleur et qu'elles introduisent le genre caractérisé par la seconde couleur. Un exemple très net en est donné avec la plante [20], dont le référent est par ailleurs inconnu (indéterminé), et qui appartient au genre *Tragopogon luteum*, dans lequel elle est en dernière position, juste avant le genre *Tragopogon purpureum* :

[20] *Tragopogon parvum ex luteo purpurascens.*

« Salsifi petit de jaune fauve tirant sur le pourpre. »

Le contexte classificatoire souligne ainsi l'iconicité de la langue (l'ordre et la place de chaque couleur est signifiant) et la valeur sémantique de chaque terme, dont la combinaison signifie systématiquement que deux couleurs sont perçues.

Un troisième contexte, d'ordre cognitif et épistémique, permet d'aller encore plus loin. Dans le cadre néo-platonicien de la Renaissance, les plantes sont conçues comme des actualisations ou des réalisations de formes idéales, à des degrés variables : une plante de couleur unie (exprimée par exemple par *luteus*) est conçue comme une actualisation parfaite de la forme idéale et constitue une plante prototypique. Inversement, une plante à plusieurs couleurs, conçue comme une actualisation imparfaite, altérée de la forme, sera éloignée du prototype. Autrement dit, dans la construction syntaxique qui nous occupe, la couleur présentée comme première est conçue comme celle de la forme et donnée comme parfaite, tandis que la couleur présentée comme seconde est conçue comme une altération par rapport à la couleur de la forme, et donc donnée comme imparfaite. Le contexte cognitif coïncide donc avec le contexte classificatoire (une couleur seconde, imparfaite, introduit le genre suivant dans lequel cette couleur est parfaitement réalisée) et avec les données linguistiques (la couleur imparfaite est exprimée par l'inchoatif en *-sc-*).

Un quatrième contexte, purement épistémique, c'est-à-dire lié à la seule configuration du savoir (Foucault 1966), achèvera notre explicitation de la construction syntaxique étudiée. Le monde des plantes est conçu à la Renaissance comme un monde de « correspondances », où toute plante établit une « correspondance » avec une autre plante ; c'est un monde fini en réseau, un monde de similitudes [21] :

[21] « [Les feuilles du Coriandre] de dessous et premier sortans, resemblent pres aux feuilles du Chervueil ou Persil, celles d'en haut & dernieres, ne sont guere dissemblables à celles là mesme, ou à celles de la Fumeterre, plus tendres toutesfois & beaucoup plus dechiquetées. » (L'Escluse 1557 : 194)

Une construction verbale paradigmatique en français l'exprime très bien, c'est « (re) tirer à » ou « tirer sur », qui signifie dans le français du XVI^e. « ressembler à », « être semblable à ». Cette construction verbale est très fréquente en botanique, comme le montrent la première attestation de « tirer sur » en français [22] et les emplois de L'Escluse dans tous les domaines [23]- [24], dont celui des couleurs [25]- [26] :

[22] « Arbre ayant les feuilles tirant sur celles de l'olivier » [Estienne 1552, article « Alaternus »]

[23] « [Les feuilles de l'Ammi] d'en haut retirent aux feuilles du Fenail ou de l'Aneth. » (L'Escluse 1557 : 192)

[24] « [Melisse] tirant à l'odeur de Citron » (L'Escluse 1557 : 182)

[25] « La fleur [de la Flambe bastarde] est bleue, tirant sur le pourpre » (L'Escluse 1557 : 143)

[26] « Les fleurs [de la Carotte sauvaige] sont blanches... au milieu desquelles se trouue vne fleurette ou deux rougeatre tirant sur le noir » (L'Escluse 1557 : 200)

Ce qu'apporte le contexte épistémique est ici essentiel : il permet de comprendre que la construction verbale « (re) tirant à/sur » introduit une deuxième couleur comme transitionnelle en ce qu'elle établit une « correspondance » avec la couleur caractéristique du genre suivant. Il se confirme, une fois encore, que c'est à travers cette deuxième couleur que s'établissent les « correspondances » entre les plantes – il s'agit toujours d'un paradigme épistémique où les choses s'appréhendent en termes de continuité et de proximité ou de ressemblance. La mise en contexte épistémique a aussi le mérite de pointer un problème, qui ressortit à la traduction en français moderne : dans cet état de langue, « tirer sur » est compris comme « tendre à », c'est-à-dire comme une nuance, et non plus comme une relation d'affinité. Avec la traduction, la dimension épistémique de la couleur risque de disparaître au profit d'une appréhension purement descriptive.

3.3. Pratiques sociales

Pour conclure sur l'utilité du contexte en matière de traduction, même dans les cas apparemment simples de couleur unique, j'aurai recours à une dernière construction [27], qui impliquera le recours à un dernier type de contexte :

[27] Advb + Adj. _[couleur]

[28] *Narcissus maximus pallide luteus.*

« Narcisse le plus grand de tous jaune fauve décoloré. »

[29] *Erica procumbens dilute purpurea.*

« Bruyère penchée pourpre délayé. »

[30] *Hyacinthus stellaris obsolete caeruleus.*

« Jacinthe étoilée bleu passé. »

[31] *Hyacinthus stellaris sature caeruleus.*

« Jacinthe étoilée bleu saturé. »

Du point de vue sociohistorique, l'expérience des couleurs est principalement liée aux pratiques tinctoriales (Pastoureau 2000). Ce qui est ici remarquable, c'est que la majorité des intensifs employés a une racine commune avec des verbes utilisés en teinturerie ou a à voir avec le domaine du vêtement. Cette dimension tinctoriale des mots est glosée en [28a]-[31a] :

[28a] *pallide* = pâli par le soleil, décoloré (≠ pâle)

[29a] *dilute* = dilué, délayé (≠ clair)

[30a] *obsolete* = passé (comme une étoffe : *obsoletus* est souvent appliqué aux vêtements usés, passés. *Homo obsoletus* = homme en haillons)

[31a] *saturate* = saturé (imprégné fortement comme un tissu) (≠ intense)

Dans ce paradigme très cohérent de mots « tinctoriaux », l'adverbe *intense* n'apparaît pas, ou du moins apparaît très tardivement (vers 1580) et reste très peu utilisé à côté de *saturate* jusqu'en 1620. Tous ces mots marquent moins le degré qu'une expérience de la couleur dégradée ou améliorée, le point médian n'étant pas un degré, mais le plaisir apporté à l'homme par une forme de couleur stabilisée, convenable, neutre – nous retrouvons là la dimension anthropocentrique évoquée plus haut (§1.1.) :

[32] *Hyacinthus stellaris caeruleus amoenus*.

« Jacinthe étoilée bleu agréable. »

[32a] *amoenus* = agréable, plaisant (≠ tendre).

Là où les traducteurs voient volontiers un degré (un bleu clair, intense, tendre...), le contexte sociohistorique apporte une sérieuse rectification. En concordance avec les données du contexte épistémique et l'idée de plantes conçues comme des actualisations d'un type ou d'une forme idéale, le degré d'intensité est d'abord perçu en termes d'altération ou d'amélioration : l'intensité de la couleur indique une réalisation idéale (absolue) ou une altération du type. Nous retrouvons encore une fois l'opposition entre réalisation parfaite et réalisation imparfaite du type. La multiplication des contextes corrobore ainsi l'explication avancée et permet même de conclure que la construction syntaxique [15c] n'a effectivement rien à voir avec celle de [27] :

[33] *luteopallescens* ≠ *pallide luteus*

La construction *luteopallescens*, comme en [18], signifie deux couleurs, la deuxième étant imparfaitement réalisée ; tandis que *pallide luteus*, comme en [28], signifie une seule couleur, elle aussi imparfaitement réalisée. L'association des différents contextes permet de voir une opposition, et non une synonymie, et de placer cette opposition dans le registre métaphysique de réalisation des Formes et non dans celui, physique, de la luminosité (jaune pâle).

4. CONCLUSIONS

Contextualiser les différentes approches constitue une aide majeure à l'interprétation sémantique et à la correction et l'exactitude des traductions. C'est une aide nécessaire lorsque les formes linguistiques, comme les chromonymes, ne sont ni sémantiquement prévisibles (exemple des adjectifs composés), ni transparentes (exemple des pseudo-intensifs), ni systématiques (exemple de la non-synonymie entre *pallide* et *pallescens*). Il n'en reste pas moins un problème : c'est que les traductions en vernaculaire sont souvent incapables de rendre la mesure de toute une épistémè et des pratiques sociales, classificatoires, etc. qui font partie de cette épistémè. On ne peut donc échapper à la nécessité d'accompagner d'un appareil critique la traduction d'une langue spécialisée, surtout quand elle est morte, pour en mettre en perspective les spécificités conceptuelles.

BIBLIOGRAPHIE

- Anonyme, trad. Fr., 1549 : Fuchs (L.), *Histoire des plantes*, Paris, veuve Arnould Byrkman.
- Bauhin (C.), 1623 : *Pinax Theatri Botanici*, Basileae, König.
- Berlin (B.), Kay (P.), 1969 : *Basic Color Terms : Their Universality and Evolution*, Berkeley, University of California Press.
- Bouveresse (J.), Rosat (J.-J.), édit., 2003 : *Philosophies de la perception. Phénoménologie, grammaire et sciences cognitives*, Paris, Odile Jacob.
- Dubois (D.), édit., 1991 : *Sémantique et cognition. Catégories, prototypes, typicalité*, Paris, CNRS.
- Estienne (R.), 1552 : *Dictionarium Latinogallicum*, 3^e édition, Lutetiae, Carolus Stephanus.
- Foucault (M.), 1966 : *Les mots et les choses*, Paris, Gallimard.
- Fuchs (L.), 2001 : *Le nouvel Herbar de 1543* (K. Dobat et W. Dressendörfer édés), Köln, Taschen.
- L'Escluse (C. de), trad. fr., 1557 : Dodoens (R.), *Histoire des plantes*, Anvers, J. Loë
- Maignan (E.), 1549 : *Commentaires tres excellents de l'hystoire des plantes composez premierelement en latin par Leonarth Fousch*, Paris, Jacques Gazeau.
- Nicot (J.), 1606 : *Thresor de la langue françoise*, Paris, Douceur.
- Pastoreau (M.), 2000 : *Bleu. Histoire d'une couleur*, Paris, Seuil.
- Sapir (E.), 1968 : *Linguistique*, Paris, Éditions de Minuit.
- Selosse (P.), [à paraître] : « Les adjectifs composés de couleur en néo-latin botanique de la Renaissance », dans *Actes du XXIV^e CILPR*, Aberystwyth.
- Touratier (C.), 1994 : *Syntaxe latine*, Louvain-la-Neuve, Peeters.

LA RECHERCHE TERMINOLOGIQUE DANS UN DIALECTE OUVERT : LE CAS DU DIOULA

Mamadou Lamine Sanogo

département linguistique et langues nationales

institut des sciences des sociétés (INSS)

Centre national de la recherche scientifique et technologique (CNRST)

Ouagadougou, Burkina Faso

« Nommer une langue, c'est certes, en partie refléter une existence autonome. Mais, c'est aussi d'une certaine manière, mobiliser au service de la dénomination le pouvoir des mots : donner des noms différents au même système, c'est déjà le morceler ; maintenir le même nom aux systèmes différents, c'est les maintenir ou les rendre unifiés ».

Marcellesi J.-B. (1981 : 61)

Résumé : *Le présent article montre que la recherche terminologique dans un dialecte ouvert doit tenir compte de la dynamique sociolinguistique des formes en présence ainsi que de leur représentation et usages chez les locuteurs. En outre, le partage des résultats et la capitalisation des recherches sont deux attitudes à observer lorsque l'on veut créer des documents ou des ressources profitables à l'ensemble des locuteurs de la langue.*

Mots-clés : *dialecte ouvert, dioula, langue véhiculaire, mandingue, bambara*

Le présent texte est inspiré des travaux de recherches terminologiques que nous menons sur le dioula véhiculaire au Burkina Faso. Ces recherches terminologiques ont été menées dans le cadre des travaux d'équipement linguistique comme la production de documents pédagogiques, la traduction, la localisation, la création de contenus multimédias... La conduite de ces travaux se fait dans l'esprit d'unité linguistique à l'instar de celui du projet MAPE (Promotion des langues manding et peul) tel qu'il avait été conduit par l'ACCT (Agence pour la coopération technique et culturelle) dans les années 1990. En d'autres termes, il s'agit de faire en sorte que les travaux de recherches terminologiques menés au Burkina Faso profitent aux équipes qui travaillent sur le mandingue au Sénégal, au Mali, en Côte d'Ivoire, en Gambie, en Mauritanie, en Guinée, au Libéria...

Or, comme nous le montrons dans Sanogo (2000) le dioula, langue des présentes recherches, est un ensemble de pratiques, un continuum dans lequel il est difficile de matérialiser les frontières. Dans ce contexte, la recherche terminologique reste confrontée à des préoccupations majeures dont les fonctions et représentations des formes en présence, l'interpénétration des formes, la pauvreté et/ou richesse de certaines formes... À cet effet, nous sommes confronté à de nombreuses difficultés dont :

- Comment gérer les rapports entre les différentes variétés dans le cadre des recherches terminologiques : quelles formes ou variétés faut-il privilégier lorsque l'on sait que

les formes ethniques du mandingue (bambara, dioula ethnique et malinké) sont « riches » mais grégaires et que les formes véhiculaires sont « pauvres » mais très dynamiques ?

- Comment choisir les termes dans une forme véhiculaire qui n'a aucune assise référentielle – au sens de Gobard (1976) – chez les locuteurs ? La situation est telle que tous les locuteurs s'y reconnaissent car ces formes restent, comme nous le montrons dans Sanogo (2002), « le parler de tous et la langue de personne » ;
- Quel sera l'impact des choix dans la gestion « des rapports conflictuels » entre les différentes variétés en présence ?

Pour mener ces réflexions, nous sommes parti de l'hypothèse selon laquelle les locuteurs sont confrontés à un conflit interne qu'est l'aspiration à exprimer leur « identité dialectale » et la tendance à clamer « l'unité linguistique » du mandingue. Au-delà de l'équipement linguistique, les travaux de recherche terminologique révèlent la dimension glottopolitique de cette entreprise.

Ces recherches présentent, pour le sociolinguiste que nous sommes, de nombreux intérêts dont les principaux sont les suivants :

- Elles permettent de jeter les bases d'une volonté de travailler désormais en partage avec ceux qui ont les mêmes intérêts et qui opèrent sur le même terrain. Il s'agit de rompre avec des démarches isolées, éphémères et sans portée importante. Les lexiques ou dictionnaires sur les différents parlers sont certes utiles, mais dès qu'il s'agit de recherches touchant à l'ensemble linguistique comme la création de logiciels, l'idéal est de penser à l'ensemble des locuteurs du mandingue ;
- Les travaux entrepris dans un espace dialectal mandingue comme le Burkina Faso sont susceptibles désormais de circuler dans les autres comme le Mali, la Gambie... S'il est connu qu'aucune équipe dans la sous-région ne dispose ni de ressources matérielles, techniques et personnelles et qu'aucune n'est capable de travailler à s'autosuffire.
- Le besoin de collaboration avait été ressenti comme une nécessité par les politiques au plan national et international puisque des travaux ont été convoqués dans le but d'harmoniser l'orthographe, le ton... à Bamako en 1966 puis à Londres en 1972. Si ces travaux sont arrêtés faute de réseau capable de les conduire, l'occasion est désormais trouvée de susciter cette nécessité de travailler en synergie afin de capitaliser, d'une part, les acquis des différentes équipes et d'autre part de briser l'isolement des chercheurs ;
- Dans des domaines assez pointus comme la localisation de ressources informatiques, notamment celle de *Openoffice*, nous n'envisageons aucunement de mettre en place un programme qui ne sera utilisable que par les locuteurs d'un seul dialecte. Nous avons lancé à cet effet un appel à tous les linguistes travaillant sur l'aire mandingue.

Inspirée de la sociolinguistique variationniste, notre méthode consiste à mener des recherches auprès des locuteurs qui utilisent les formes différentes. Notre souci d'objectivité nous amène à ne pas seulement nous contenter des formes urbaines dont nous sommes nous-

même locuteur natif, mais à faire en sorte que toute la langue mandingue soit réellement représentée dans l'analyse. Ainsi, nous consultons des locuteurs urbains, ruraux, des locuteurs d'autres variétés très distantes comme le dafin, le bobo, le dioula ethnique de Côte d'Ivoire, le bambara du Mali, le malinké de Guinée... Nous consultons également les dictionnaires et lexiques élaborés à partir d'autres dialectes.

Le présent texte comporte trois parties : dans un premier temps, nous allons présenter la situation sociolinguistique du dioula véhiculaire. Ensuite, nous montrerons les difficultés liées à la nature du dioula en tant que dialecte ouvert et enfin, nous allons explorer quelques pistes de solutions que nous avons prospectées au cours de travaux menés au Burkina Faso.

1. DE LA NATURE DU DIOULA VÉHICULAIRE

Groupe de dialectes de la branche est de la langue mandingue, le dioula est un ensemble dialectal complexe et dynamique. Il constitue, depuis les travaux de Delafosse (1929), avec le malinké et le bambara les trois grands groupes de dialectes du mandingue. Il représente sans doute le groupe aux affinités linguistiques les moins décrites ; ses contours géographiques restent à dessiner car les formes ou variétés entrant dans sa composition sont insaisissables. Le dioula est, à l'instar des autres parlars à fonction véhiculaire des grandes villes africaines, un ensemble convergent, composite et hétérogène où se mêlent des formes linguistiques le plus souvent apparentées, le tout dans un processus de glottogenèse permanent. Cet aspect sociolinguistique, très souvent non perçu dans les études de description systématique laisse voir des nombreuses variétés dialectales coexistantes et juxtaposées. Cependant, la réalité sociolinguistique du dioula est beaucoup plus complexe, compte tenu de la présence quasi simultanée de formes plus ou moins distinctes sur le même espace urbain, mais caractérisée chacune par sa dynamique propre. Ainsi, il faut garder à l'esprit que nous sommes en présence d'un continuum linguistique ou dialecte ouvert – au sens de Hendersen (cité par Calvet 2000) – au sein duquel des formes plus ou moins apparentées se rencontrent, se frottent et se mélangent.

Des linguistes ayant travaillé sur la langue mandingue – dont le dioula est une variante – en contexte urbain en Côte-d'Ivoire (Tera (1986), et Partmann (1975), au Burkina Faso (Coulbaly (1984), Coulbaly (1983), Sanogo (1996, 2000) et au Mali (Canut 1996), ont perçu ces particularités dans les zones urbaines. Il ressort des différents travaux sur les variantes urbaines des pratiques langagières mouvantes et instables tendant vers la disparition – par extinction – des traits locaux ou particularités régionales, notamment chez les jeunes. N'est-ce pas ce qui a fait croire à Keita (1990) qu'il existe un dioula urbain vernacularisé à Bobo-Dioulasso ?

En plus d'être le fruit de cette dynamique des formes du mandingue, le dioula véhiculaire comporte de nombreux emprunts au français, langue de l'administration. Cette interférence n'est pas seulement lexicale mais aussi phonologique et même souvent grammaticale. En retour, ce dioula contribue également à enrichir le français africain et cela se ressent surtout dans le français populaire ivoirien (FPI) où les apports sont très importants. Toute cette dynamique sociale se déroule dans un contexte marqué par un substrat d'arabe qui a été la langue de la religion musulmane chez les Dioula. Le système comptable du dioula, le calendrier ainsi que le vocabulaire du droit, de l'administration, de la justice... restent profon-

dément marqués par cette langue dont le domaine d'usage est réduit à ce jour aux salles de classe.

Pour conclure sur la nature sociolinguistique du dioula, disons qu'à l'instar des « dialectes ouverts », le dioula véhiculaire comporte un centre formé par les tendances urbaines (dioula commercial et dioula urbain) assurant la fonction véhiculaire et des périphériques (dioula ethnique, bambara, etc.) en voie de fusion. Les formes urbaines alors issues de ce processus de glottogenèse permanent comportent des apports multiples dont le nombre varie en fonction des locuteurs.

2. DES DIFFICULTÉS DE RECHERCHES TERMINOLOGIQUES

Les conflits consécutifs à la coexistence des formes différentes du mandingue trouvent leur terrain privilégié d'expression dans le cadre de la recherche terminologique. En effet, dès qu'il s'agit de recenser les mots ou termes d'un domaine spécifique comme l'administration, le droit, la technologie, la pédagogie... les informateurs avec qui nous travaillons ont deux attitudes :

- Premièrement, ils tentent de mieux structurer leurs pratiques. Au lieu de retrouver les termes ou réalisations phonologiques dans leurs usages de tous les jours, ils se mettent à les rechercher dans « un parler pur ». C'est l'occasion idéale de les voir se ressourcer soit dans leurs formes ethniques (bambara et dioula ethnique) « chez nous on dit » ou soit dans des formes prestigieuses (bambara) « c'est comme ça dans la langue pure ».
- Deuxièmement, ils tentent de justifier leurs pratiques hétérogènes en « montrant leurs limites » de la maîtrise de la langue dioula. Ainsi, ils justifient leurs attitudes par des phrases du genre « Comme ce n'est pas notre langue, nous parlons comme ça ». Par conséquent, leurs pratiques restent marquées par des influences des formes multiples ainsi que le français et autres langues africaines. Nous retrouvons chez les locuteurs qui justifient cette deuxième attitude des interférences de toute sorte (phonologiques, grammaticales, lexicales...).

Dans le cadre des travaux sur le lexique des droits humains et du titre I de la constitution du Burkina Faso, nous avons organisé deux ateliers aux cours desquels nous avons consulté des « professionnels » (formateurs, juristes, politologues, traducteurs, chercheurs, communicateurs...) qui travaillent sur le dioula au Burkina Faso. En plus d'être confrontés, de par leur travail au sein de la Sous-commission nationale du dioula¹, aux usagers des autres dialectes de la langue, ils représentent la quasi-totalité des locuteurs des formes du dioula en présence au Burkina Faso.

Au cours de ces travaux, nous nous sommes rendu compte que la variation n'était pas une préoccupation de sociolinguiste uniquement mais, un problème auquel terminologue et lexicographe allaient être confrontés. En effet, le choix des termes dans les différentes for-

¹ Les équipes qui travaillent sur les langues du Burkina sont organisées au sein d'une structure centrale dénommée Commission nationale des langues burkinabè. Cette dernière est organisée en Sous-commission nationale pour chaque langue.

mes ou dialectes a été la première difficulté. Cette question avait ainsi une façade onomasiologique et une autre sémasiologique.

Ainsi, au plan onomasiologique, comment choisir par exemple entre deux mots relevant de formes sociologiques différentes, quand on sait que l'agitateur est dit *lasulikèla*² en dioula véhiculaire alors qu'en bambara on dirait *gololamagala*. Si le contexte permet de faire la différence entre ces deux mots relevant chacun d'un niveau de langue différent, c'est dans la maîtrise de la culture du milieu que le terme bambara trouve tout son sens. De même, lorsque dans le domaine du commerce, le terme le plus utilisé est *feerefèn* (chose + vente) du bambara, c'est le dioula de Kong qui nous donne un terme de la langue mandingue *karijogo*. Cependant, ce dernier terme n'a pas été retenu par les membres de la commission, prétextant qu'il n'est pas connu et utilisé comme le néologisme bambara. Toujours sur le plan phonologique, certaines nuances minimales comme des variantes libres *d/l* dans *la* (dioula) et *da* (bambara) « se coucher » ne passent pas inaperçues. Ainsi, les membres de la commission n'ont pas hésité à retenir *lalôniya* au lieu de *ladôniya* qui a l'occurrence la plus élevée. Il en est de même pour la nasalisation et la dénasalisation des voyelles du dérivatif d'attribution à une localité *kan/ka*. Il a ainsi été retenu *Bobokan* au lieu de *Boboka* « un Bobolais ».

En outre, sur le plan sémasiologique, les termes communs aux formes différentes ayant des valeurs sémantiques différentes posent également un réel problème de choix. Par exemple, si *bolon* veut dire « piquet » en bambara, en dioula véhiculaire ou en dioula de Kong, ce mot désigne un groupe ethnique proche des *Marka* (Diallo 2000). Par conséquent, son usage dans le domaine de l'informatique pour « ancre » nécessitera une explication fondée sur son sens en bambara. Il en est de même pour *burun* qui désigne « le cylindre » en bambara. Il sera préféré au même mot désignant « danse (spécifique) » dans le dioula de Kong.

Pour ce qui est des termes du droit, de la gouvernance, de la démocratie, de la justice etc., nous avons constaté deux tendances générales chez les informateurs et nos collègues linguistes.

La première consiste à emprunter des termes au français qui est la langue officielle. L'usage courant ayant consacré ces emprunts, de nombreux membres des commissions ont proposé qu'ils soient maintenus en donnant les termes dans la langue nationale s'il y en a. Le paradoxe de cette décision est qu'ils ont également proposé que dans les entrées, le terme en langue nationale soit prioritaire, pour des « soucis de valorisation » de ces langues. En revanche, lorsque le terme n'existe pas en langue nationale, l'emprunt seul demeure à l'entrée.

La deuxième attitude consiste à rechercher dans le fonds culturel ancien des termes qu'évoque le référent en français. Ainsi, dans le domaine du droit et de la justice, par exemple, les termes dits « authentiques » que les membres se sont empressés de proposer se sont révélés être des termes arabes. C'est le cas du juge *kadi*, du droit *sharia*...

² Nous avons décidé de marquer le *e* ouvert par *è* ainsi que le *o* ouvert par *ô*. Par conséquent, nous n'écrivons pas les tons.

Quant aux travaux de localisation et de création de contenus multimédias, nous sommes confrontés, à l'instar des autres domaines, à la variation. Notre démarche de recherche terminologique consiste à consulter en priorité le MAPE en tenant compte du contexte. Ainsi, les termes que nous trouvons pour *caractère*, *arrière-plan* et *police*, *adapter*, *annuler*... ne conviennent aucunement dans le domaine du numérique. En effet, ce domaine spécifique nécessite une approche ayant en vue une dimension géographique plus large. Car si un dictionnaire ou un lexique peut être utilisé au niveau d'un seul état ou d'une portion linguistique restreinte, une plate-forme ou un logiciel en langue nationale devra être élaborée dans une perspective de diffusion et de circulation sur un espace plus important. Ainsi, si *Openoffice*, qui est un logiciel de bureautique libre, doit être localisé dans cette langue, il faut avoir à l'esprit que cette entreprise devra se faire au bénéfice de la communauté des locuteurs du mandingue. En rupture avec ce qui a été fait dans le cadre du programme *Word* par l'équipe de Dakar, nous pensons que l'on gagnerait à localiser au profit d'un groupe plus large indépendamment des frontières politiques ou associatives.

En somme, se présentent ici quelques grands problèmes que nous avons rencontrés dans le cadre des recherches terminologiques, à cause de l'extrême variabilité du dioula. Il convient de signaler le manque de projet de standardisation. Progressivement et sans qu'il y ait un texte ou une directive quelconque, nous avons constaté que les locuteurs accordent plus de prestige au bambara même s'ils ne le parlent pas. Sans doute lié à son prestige historique remontant à l'empire du Ghana puis du Mali avec le commerce transsaharien, cette langue va renforcer son assise avec les États dioula du Guiriko, de Kong ainsi que les puissants royaumes Bambara de Ségou et du Kartha (Mage 1980). Plus récemment, les rôles et fonctions des interprètes, assurés sous l'administration coloniale (langue de commandement des troupes de « tirailleurs sénégalais »)... ont largement contribué au prestige de cette portion du mandingue.

Cependant, toutes les difficultés que nous venons de soulever ne sont pas insurmontables. Le contexte oral défavorable à la circulation des termes consacrés par les dictionnaires impose de l'imagination face à l'ampleur des tâches.

3. DES PISTES DE SOLUTION

Nous avons ainsi envisagé un certain nombre de solutions pour faire face aux contraintes liées à la nature du dialecte ouvert. À cet effet, les plus importantes sont :

La première consiste à mieux analyser la nature des documents et ressources à élaborer, car nous sommes dans une phase où les habitudes d'exploitation de documents écrits ne sont pas bien ancrées. Ainsi, il convient de faire la distinction entre les documents et ressources entrant dans le cadre de l'apprentissage et ceux destinés à l'exploitation chez un public maîtrisant la lecture et l'écriture.

En effet, les documents et ressources destinés à l'apprentissage sont surtout les documents pédagogiques destinés aux enfants dans les écoles bilingues, aux alphabétisés et néoalphabétisés désirant renforcer leurs acquis...

Nous avons estimé, au vu des difficultés que cela présentait, qu'il faut minimiser la question de la variation à ce niveau. Il ne s'agit pas de l'occulter mais de faire en sorte qu'elle ne constitue pas une source supplémentaire de difficultés dans l'apprentissage.

Ainsi, nous avons remarqué dans le cadre de ces recherches que le recours aux dérivatifs comme source de création des néologismes est très accentué alors que nous avons démontré la baisse de fréquence de ce procédé dans les formes véhiculaires (Sanogo 1996). De même, les séquences médiates du syntagme complétif ainsi que la fréquence des occlusives vélares doubles (*kp* ou *gb*) sont considérées comme des survivances grammaticales et phonologiques du dioula ethnique, une forme référentielle mais moins prestigieuse ; leur usage devra fait l'objet de surveillance.

Cependant, il convient, pour les documents pédagogiques comme les livres pour enfants, les ressources d'alpha... d'éviter de faire de l'apprentissage de l'écriture une épreuve de glottopolitique. En effet, si les enfants sont sensibles aux formes variables, les adultes le sont davantage. Dès qu'ils sentent que l'apprentissage tend à enraciner chez eux une forme prestigieuse de la langue, ils sont motivés. En revanche, lorsque l'instructeur parle une forme moins prestigieuse avec des emprunts à d'autres langues minoritaires comme le dafin, le bwabu, ils en concluent que celui-ci « parle un mauvais dioula ». Nous avons testé leur réaction à partir des pratiques langagières chez certains animateurs de radio ou de télévision en dioula. Ceux qui utilisent les formes proches du bambara semblent avoir plus de succès auprès des auditeurs.

Quant aux documents et ressources destinés à l'exploitation, nous rangeons dans cette catégorie les traductions des textes du système des Nations unies, les conventions et protocoles internationaux, les constitutions des États, des lexiques de santé, les plateformes multimédias, les logiciels...

Pour les documents et ressources devant circuler chez ceux qui ont des habitudes de lecture, nous proposons de s'inspirer de l'expérience du MAPE. Cette expérience qui a consisté à explorer des domaines aussi divers comme la pédagogie, la linguistique, les mathématiques, l'histoire, la géographie, le droit, les sciences de la vie et de la terre... a proposé des termes harmonisés. L'objectif étant de faciliter la circulation des termes et des documents entre les différentes sphères, ces termes sont donc valables au Mali, au Burkina Faso, en Côte d'Ivoire, en Guinée, au Sénégal, en Mauritanie...

La démarche que nous proposons n'est pas nouvelle. Elle consiste juste à rappeler des principes ayant fait leurs preuves, principes qui sont en abandon au profit des initiatives isolées. Au-delà de l'espace francophone dont les frontières ne correspondent en rien à celles du mandingue, nous envisageons une étroite collaboration entre manding (anglophones) et mandingue (francophone). Il s'agit de briser cette frontière artificielle afin de profiter des ressources de part et d'autre. Nous sommes conscient des difficultés supplémentaires, compte tenu des emprunts à l'anglais ou au français de part et d'autre, mais nous estimons que le fonds commun mandingue est plus important que cette partie du vocabulaire importé.

Par conséquent, nous proposons de prendre le lexique du MAPE comme base, quitte à l'amender au vu des résultats des recherches. Conscient que ce document a des limites,

compte tenu de son âge (20 ans) et de l'évolution des domaines d'usage qui ont été étudiés au cours de son élaboration, il faut envisager une adaptation à travers une forme dynamique à mettre en ligne, si possible.

Les avantages de cette innovation sont nombreux :

Premièrement, la mise en place d'un dictionnaire électronique sous la surveillance d'une équipe technique permettra de disposer de ressources linguistiques à coût moindre et accessible.

Ensuite, le caractère dynamique et interactif fait qu'il sera alimenté progressivement grâce aux utilisateurs et aux chercheurs qui seront amenés à le consulter régulièrement et à déposer leurs arguments et produits à travers les forums.

Enfin, la mise à disposition de cette ressource permet de gagner du temps dans les traductions et la création de ressources.

Cependant, un certain nombre de précautions doivent être prises afin de faciliter d'une part cette mise en ligne et d'autre part, l'exploitation des données.

Ainsi, les formes de création par néologisme comme la dérivation, la composition devront s'inspirer de la dynamique sociolinguistique. Si des formes grégaires accordent plus d'attention à la dérivation, les études menées en Côte-d'Ivoire (Dérive 1975) et au Burkina Faso (Sanogo 2000) montrent la préférence des locuteurs pour les formes non dérivées. De même, la composition étant très liée au système tonal (compacité tonale), les formes composées devront être revues puisque dans nombre de pays, il a été décidé de ne pas marquer le ton dans l'orthographe.

En conclusion, disons que si les dialectes ouverts sont des espaces privilégiés de questions glottopolitiques, la recherche terminologique est une occasion de mettre en exergue des pratiques et attitudes des locuteurs et usagers de la langue. De la gestion des formes prestigieuses aux choix des emprunts et autres génies propres de la langue, le terminologue est interpellé à chaque instant de ses recherches.

Il doit fournir des réponses à chaque interrogation sur les raisons de telle ou telle stratégie car au-delà même de l'élaboration d'un document ou d'une ressource, il guide les usages, recommande les pratiques qui vont s'installer progressivement. L'élaboration d'un lexique ou d'un dictionnaire n'a rien d'une tâche scientifiquement neutre. Les entrées vedettes sont des choix politiques et la volonté de mettre des variantes est une décision de politique linguistique.

Les ressources informatiques, les contenus multimédias qui sont destinés à un public plus large et peut-être plus dynamique et plus apte au changement doivent certes intégrer le génie de la langue, mais aussi se coller à leur contexte. C'est donc l'occasion d'ouvrir davantage vers les emprunts comme le cas du français aux anglicismes de l'informatique.

Comme nous le voyons, la recherche terminologique, à ce stade, nécessite une certaine organisation, une philosophie de la mémoire, de la capitalisation des acquis en rupture avec des élans de traduction spontanée. La mise en place de ressources en ligne ou sur papier devra guider désormais toute démarche en la matière car non seulement les usagers, locu-

teurs et chercheurs devront profiter de ce qui est fait, mais ils devront trouver également un espace où ils déposeront les fruits de leurs travaux. MAPE a été un exemple et une expérience ayant produit des fruits, il nous appartient d'imaginer des perspectives nouvelles intégrant nos potentialités et les opportunités technologiques qui sont à notre portée.

BIBLIOGRAPHIE

- ACCT, 1983 : Projet MAPE, structures et méthodes du projet de promotion des langues manding-peul, Documentation française, Paris.
- Calvet (L.-J.) et Moussirou-Mouyama (A.), 2000 : *Le Plurilinguisme urbain*, OIF, Institut de la Francophonie, Québec, Canada.
- Canut (C.), 1996 : « Instabilité des usages et non fermeté du système manding du Mali », *Mandenkan*, n° 31, *Bulletin semestriel d'études linguistiques mandé*, INALCO, CNRS, Paris, pp. 76-91.
- Coulibaly (B.), 1984 : *Le jula véhiculaire de Haute-Volta*, thèse d'État, Université de Paris V, vol. I-II.
- Delafosse (M.), 1929 : *La langue mandingue et ses dialectes*, Librairie orientaliste Paul Geuthner, Paris.
- Derive (M.-J.), 1975 : « Dioula véhiculaire, dioula de Kong et dioula d'Odienné » dans les *Annales de l'Université d'Abidjan*, série H, fascicule I, pp. 55-84.
- Diallo (M.), 2000 : « Traits spécifiques au marka dans l'ensemble dialectal mandingue », dans *Cahiers du CERLESHS*, n° 2 spécial 2000, Université de Ouagadougou, pp. 13-25.
- Galtier (G.), 1980 : *Problème dialectologique et phonographématique des parlers mandingues* ; thèse pour le doctorat de troisième cycle, Université Paris VII.
- Gobard (H.), 1976 : *L'Aliénation linguistique. Analyse tétraglossique*, Flammarion, Paris.
- Keita (A.), 1990 : *Esquisse d'une analyse ethno-sémasiologique du jula vernacularisé de Bobo-Dioulasso*, thèse pour le doctorat nouveau régime, Université de Nice.
- Mage (E.), 1980 : *Voyage dans le Soudan occidental (1863-1866)*, (rééd.), Karthala, Paris.
- Marcellesi (J.-B.), 1981 : « Bilinguisme et Diglossie » dans *Langages* n° 61, Larousse, Paris, pp. 2-127.
- Partmann (G.), 1975 : « Quelques remarques sur le jula véhiculaire en Côte-d'Ivoire » dans *Annales de l'Université d'Abidjan*, ILA, série II, 1975, pp. 241-260.
- Sanogo (M. L.), 1996 : *Langues nationales, langues véhiculaires, langue officielle et glottopolitique : La problématique du jula véhiculaire au Burkina Faso*, thèse pour le doctorat nouveau régime, Ura-CNRS 1164, Université de Rouen.
- Sanogo (M. L.), 2000 : « À propos de jula à Bobo-Dioulasso », dans les *Cahiers du CERLESHS*, 2^e numéro spécial, Université de Ouagadougou, pp. 73-83.
- Sanogo (M. L.), 2002 : « Dynamique langagière et pratique fluctuante du dioula au Burkina Faso », dans *Sciences et technique, Série Sciences sociales et humaines*, CNRST Ouagadougou, vol. XXII, n° 1, pp. 173-193.
- Sanogo (M. L.), 2002 : *Élaboration d'un lexique des termes juridiques, administratif, de bonne gouvernance et des droits de l'homme en dioula*, Secrétariat d'état aux Droits de l'homme et Programme d'appui à l'administration judiciaire, Ouagadougou.
- Sanogo (M. L.), 2003 : « L'ethnisme jula : origines et évolution d'un groupe ethnolinguistique dans la boucle du Niger », dans *Burkina Faso, Cent ans d'histoire, 1895-1995*, t. I, Paris, Karthala, pp. 370-379.
- Tera (K.), 1986 : « Le dioula de Côte-d'Ivoire, expansion et développement » dans *Cahiers ivoiriens de recherches linguistiques* n° 20, Abidjan, pp. 5-31.
- Touré (S.), 1986 : « Notion de mesure dans le langage scientifique dioula » dans *Cahiers ivoiriens de recherches linguistiques* n° 19, Abidjan, pp. 59-68.

LE TERME *BIEN COMMUN* ET LA CONSTRUCTION DU SENS. MAIS DANS QUEL CONTEXTE ?

Dorota Śliwa

institut de philologie romane
Université catholique de Lublin im. Jana Pawła II
Lublin, Pologne
dorotea@kul.lublin.pl

Résumé : *l'objet d'analyse est le terme bien commun, unité polylexicale, dont le sens est construit en discours de manière complexe. En un premier lieu, l'analyse de cette unité est opérée selon la méthode distributionnelle suivant l'analyse de la catégorie référentielle et les prédicats associés, et elle est suivie d'une étude sémantique des lexèmes synonymes relevés de la même catégorie référentielle. Ces analyses font voir que cette unité polylexicale a un statut différent, collocation ou nom composé, en fonction du discours où elle devient une formule. L'article se termine par la proposition de méthode de recherche relevant de la sémantique référentielle et cognitive.*

Mots-clés : *prédicats, collocations, discours, référence, sémantique, formule*

Le terme *bien commun*, qui est une unité polylexicale très fréquemment employée dans divers discours, intrigue le linguiste pour différentes raisons : comment se construit le sens de cette unité ? Quelle catégorie référentielle vise-t-elle ? etc.

Si nous regardons les sens des éléments constituant cette unité polylexicale, nous constatons que le N (*bien*) renvoie par métonymie à une « chose » : « valeur positive » – « chose » ; l'adjectif *commun* peut être glosé comme « appartenant à une communauté ». Le sens déduit de l'analyse morphologique et sémantique est « une chose à laquelle on attribue une valeur positive et qui appartient à quelqu'un ».

Cette analyse nous signale que pour étudier la construction du sens de cette unité polylexicale, il est nécessaire de préciser deux dimensions : la dimension référentielle des deux arguments (*chose* et *quelqu'un*) et la dimension relationnelle de ces arguments exprimée par un prédicat verbal. Pour ce faire, nous analyserons les unités lexicales relevées dans deux discours plus ou moins rapprochés par la longueur (20 000 signes environ) et par le public auquel ils sont adressés (instances internationales). Il s'agit du discours de Philippe Quéau (PQ) prononcé en 1999 dans le cadre de l'Unesco (http://2100.org/conf_queau1.html) et du discours de Jean-Paul II (JPII) pour la journée mondiale de la Paix 2005. (http://www.vatican.va/holy_father/john_paul_ii/messages/peace/documents/hf_jp-ii_mes_20041216_xxxviii-world-day-for-peace_fr.html).

En un premier temps nous procéderons à une analyse selon la méthode distributionnelle en repérant d'abord les énoncés définitoires de l'unité *bien commun* et les prédicats qui y sont

associés. Ensuite nous élargirons notre analyse en mettant l'unité *bien commun* en rapport paradigmatique avec des unités synonymiques. Nous étudierons en particulier les catégories référentielles désignées par ces unités lexicales.

1. ANALYSE SELON LES PRINCIPES DISTRIBUTIONNELS – DÉMARCHE COMPUTATIONNELLE STANDARD : RELEVER LES ÉNONCÉS DÉFINITOIRES ET LES COLLOCATIONS

Le traitement automatique du texte a pour fondement habituel la méthode d'analyse distributionnelle. Nous relevons donc les énoncés définitoires pour déterminer la catégorie référentielle et les constructions verbo-nominales ou adjectivo-nominales pour obtenir une liste de prédicats associés et analyser la construction du sens de l'expression *bien commun* en discours.

1.1. Analyse des énoncés définitoires

Les définitions produites par les deux auteurs sont formulées en discours pour expliciter le sens de l'expression *bien commun*. Leurs énoncés définitoires sont introduits par des marqueurs (formes linguistiques), repérables par l'extraction automatique et répertoriés par J. Rebeyrolle (2000) comme énoncés définitoires directs. Parmi ces énoncés nous avons repéré notamment le marqueur *entendre* qui est synonyme du verbe *signifier* et qui établit une relation de signification entre les syntagmes nominaux dans la structure syntaxique de type A V *entendre* B.

Philippe Quéau, en définissant dans son discours le *bien commun*, se réfère au droit international :

[1] Il faut aussi définir ce que l'on entend par « bien commun ». En droit international, on l'entend au sens de *res nullius* (une chose qui n'appartient à personne, comme la haute mer) ou de *res communis* (une chose qui appartient à tous, comme le spectre des fréquences électromagnétiques). La catégorie de *res publica* en tant qu'elle s'oppose à la *res privata* peut aussi s'appliquer.

Le sens de l'expression *bien commun* est défini essentiellement par la relation d'appartenance d'une « chose » à un/des « propriétaire/s », exprimée par le prédicat verbal *appartenir à quelqu'un*.

Jean-Paul II se rapporte à la définition donnée par une autre instance internationale qu'est le Vatican II :

[2] Selon une large acception, on entend par bien commun « l'ensemble des conditions de la vie sociale qui permettent aussi bien aux groupes qu'à chacun de leurs membres d'atteindre leur perfection d'une façon plus plénière et plus aisée » (Conc. œcum. Vat. II, *Const. past. sur l'Église dans le monde de ce temps, Gaudium et spes*, n° 26).

Nous ne retrouvons plus la relation d'appartenance entre une « chose » et les « propriétaires » mais une relation de potentialité entre les « conditions » et une « société » ou une « personne », exprimée par le prédicat verbal modal *permettre à quelqu'un de faire quelque chose* et le prédicat verbal désignant une action *atteindre sa perfection*.

Comparons à présent les hyperonymes relevés de ces énoncés définitoires et les référents des catégories référentielles désignées par ces hyperonymes. Dans la première définition [1], l'hyperonyme est un mot simple *chose*, qui désigne une catégorie référentielle générale. La définition ne demande pas un commentaire. L'hyperonyme indique une catégorie des référents spécifiée seulement par le rapport d'appartenance. L'auteur indique deux référents : la haute mer quand il donne le sens de *res nullius* et le spectre des fréquences électromagnétiques quand il donne le sens de *res communis*. Ces référents sont répétés par le marqueur *comme* qui tient la place du prédicat verbal *énumérer* (*une chose qui n'appartient à personne, comme la haute mer ; une chose qui appartient à tous, comme le spectre des fréquences électromagnétiques*). Dans la définition [2], l'hyperonyme *conditions* est déterminé : *l'ensemble des conditions de la vie sociale* et désigne une catégorie référentielle potentielle dont les référents sont à construire. L'auteur ne l'illustre par aucun référent.

L'analyse des énoncés définitoires relevés dans les deux discours nous conduit à constater que les auteurs, utilisant la même expression *bien commun*, parlent de deux choses différentes.

1.2. Analyse des collocations

Deuxième démarche, dans la continuité du traitement automatique des langues, est d'opérer l'extraction des constructions syntaxiques communément appelées collocations qui ont pour base l'expression *bien commun*. Si nous admettons que les collocations sont des unités pour lesquelles il est possible d'isoler plusieurs niveaux d'analyse, nous retrouvons alors à un niveau sous-jacent des « prédicats appropriés » dont il est question dans le modèle des « classes d'objets »¹. Dans les discours analysés, la majorité des collocations sont des constructions verbo-nominales par rapport au prédicat verbal et des constructions nomino-adjectivales.

Regardons d'abord les constructions verbo-nominales que nous allons étudier en fonction de la place de l'expression *bien commun* par rapport au prédicat verbal (V).

Dans le discours de PQ nous avons relevé trois collocations où l'expression *bien commun* est en position de complément d'objet, à savoir [3] (créer, gérer, respecter) le bien commun. Le sens de ces prédicats est compatible avec le sens de l'hyperonyme *chose* et fournit des informations concernant l'activité d'un agent humain par rapport à une chose.

Dans le discours de JP II les collocations sont plus nombreuses et se répartissent en deux groupes selon la position de l'expression *bien commun* par rapport aux prédicats verbaux. En ce qui concerne la construction V *bien commun*, nous avons deux groupes, selon que l'agent est : < humain > [4] (cultiver, réaliser, s'engager pour) le bien commun ou < non humain > [5] transformer le bien commun en quelque chose, dépouiller le bien commun de quelque chose². Nous constatons toutefois que seul le prédicat *réaliser* peut être compatible

¹ Il s'agit du modèle présenté par A. Clas et P. Lerat (1997), élaboré déjà par J. Dubois et développé de manière systématique par G. Gross pour le rendre « plus productif dans une analyse automatique ».

² Ces prédicats ont été extraits du contexte suivant: « des visions résolument réductrices de la réalité humaine transforment le bien commun en un simple bien-être socio-économique, privé de toute finalité transcendante, et le dépouillent de sa plus profonde raison d'être. »

avec le sens « conditions de la vie sociale ». Si donc d'autres prédicats ne sont pas compatibles avec le sens déduit de l'énoncé définitoire, où chercher alors cette compatibilité ? Nous répondrons à cette question ultérieurement. Quant à la construction *bien commun V* nous avons les prédicats *concerner*, *exiger*, *revêtir*, *prendre* mais nous sommes consciente que seule l'énumération des formes verbales nous amène à conclure que ces prédicats ne sont pas compatibles avec le sens et qu'il est nécessaire de les réinsérer dans leur contexte discursif :

[6] « La personne peut-elle donc se réaliser pleinement en faisant abstraction de sa nature sociale, c'est-à-dire de son être "avec" et "pour" les autres ? Le bien commun la concerne de près. »

[7] « Le bien commun exige donc le respect et la promotion de la personne et de ses droits fondamentaux »

[8] « Le bien commun [...] revêt aussi une dimension transcendante... »

[9] « Le bien commun [...] prend aujourd'hui une dimension de plus en plus universelle et, par conséquent, il comporte des droits et des devoirs qui regardent tout le genre humain. »

L'unité polylexicale *bien commun* a donc un sens spécifique pour chaque énoncé suivant le contexte dans lequel elle est insérée.

Passons à présent aux constructions nomino-adjectivales, peu nombreuses dans les discours analysés, c'est pourquoi nous les présenterons selon le critère de trait sémantique < dimension > et < modalité > associé au prédicat adjectival. Le trait < dimension > est exprimé par l'adjectif *mondial* chez PQ (*le bien commun mondial*) et par l'adjectif *universel* chez JPII (*le bien commun universel*). PQ développe le sens de l'adjectif (qui est un adjectif de relation construit sur la base nominale *monde*) et aboutit à la conclusion que la notion du « monde » est relative en fonction des bénéficiaires et en fonction du bien concerné³, c'est pourquoi il précise :

[10] « La notion de "bien commun mondial" vise l'intérêt collectif de tous les habitants du monde »

et

[11] « Le bien commun mondial pourrait être défini comme une "res publica" mondiale, s'il y avait une forme de gouvernance capable de la définir et de la défendre comme telle. »

³ « Mais il y a plusieurs manières de définir le "monde". Par exemple la protection de la couche d'ozone intéresse évidemment toute la population mondiale actuelle et pas seulement ceux d'entre nous qui sont les plus exposés aux rayonnements nocifs. En revanche les manipulations du génome humain concernent aussi l'avenir de l'espèce humaine et pas seulement les générations contemporaines. Les questions liées à Internet touchent directement les 2 % d'internautes de la planète, mais affectent indirectement les info-exclus... ».

JPII n'explique pas le sens de l'adjectif *universel* qui semble ne pas nécessiter de précisions supplémentaires. Le deuxième trait < modalité > est associé à l'adjectif *vrai* employé par JPII et dont le sens est expliqué en contexte discursif :

[12] « [...] les chrétiens savent que Jésus a mis en pleine lumière la réalisation du vrai bien commun de l'humanité. »

Remarquons enfin que le sens du prédicat adjectival est construit non pas par rapport aux propriétés inhérentes de « chose » ou de « conditions de la vie sociale » mais par rapport à une relation entre ce qui est désigné par l'expression *bien commun* et les autres entités impliquées par le discours. Une fois de plus, il serait difficile d'appliquer automatiquement la règle de compatibilité de traits sémantiques du N et du prédicat adjectival.

En regardant la construction du sens selon les principes distributionnels, nous constatons que le sens du *bien commun* échappe dans le discours de JPII et que l'approche distributionnelle, purement formelle, sans prendre en compte le réel désigné par les unités lexicales, se révèle insuffisante pour décrire la construction du sens en contexte qu'est le discours. Que faut-il faire alors ?

2. LA DIMENSION RÉFÉRENTIELLE DE L'EXPRESSION *BIEN COMMUN* ET DE SES SYNONYMES

Tout en travaillant sur les unités lexicales repérables par le traitement automatique, nous proposons de chercher des unités lexicales qui sont en rapport paradigmatique avec les unités constitutives, à savoir : *bien* et *commun*, et d'étudier les référents désignés par ces unités lexicales (*choses, créateur, propriétaire, administrateur, bénéficiaire*).

2.1.

Regardons d'abord la présence du substantif *bien* tout seul et sa définition dans les discours étudiés et les adjectifs qui l'accompagnent ; ensuite, les référents désignés par le N et quels indices discursifs les signalent.

Dans le discours de Philippe Quéau, nous n'avons pas retrouvé le N *bien* isolé, sauf un SN *biens appartenant au domaine public* où l'expansion de N correspond à la paraphrase définitionnelle. Apparemment, la catégorisation des référents désignés par ce lexème n'est pas l'objet du discours, elle est tout de même implicite car nous avons relevé le SN *biens immatériels*, ce qui nous autorise à déduire que la seule catégorisation référentielle opérée par PQ concerne la distinction entre *choses matérielles* et *choses immatérielles*.

Examinant sur le plan paradigmatique les unités polylexicales à trois éléments parmi lesquels l'adjectif *mondial* se retrouve toujours, nous avons constaté l'existence des unités polylexicales synonymiques : *bien commun mondial*, « *res publica* » *mondiale*, « *chose publique* » *mondiale*, *biens collectifs mondiaux* (« *res communis* », « *res nullius* » et « *res publica* »). Nous avons affaire à la variation synonymique de type lexème de l'expression définie et lexème de la glose définitionnelle (cité d'ailleurs par les guillemets), entre N (*bien* / « *res* » / « *chose* ») et A (*commun* / « *publique* » / *collectif* (*communis* », « *nullius* » « *pu-*

blica »)). L'auteur emploie ces synonymes pour désigner le même ensemble de référents, comme le prouvent les indices de type : a) verbe *citer* :

[13] « Les exemples de biens collectifs sont nombreux : citons pêle-mêle la couche d'ozone... les fréquences hertziennes ... »

b) ou les parenthèses :

[14] « [...] les riches peuvent avoir un avantage relatif plus important à tirer de certains biens communs (haute mer, espace, fréquences hertziennes) que les pauvres »

c) ou encore par la préposition *comme* :

[15] « La gestion des biens communs de l'humanité (comme l'eau, l'espace, le génome humain,... mais aussi le patrimoine culturel public, les informations dites du "domaine public", les idées, les faits bruts) doit désormais être traitée comme un sujet politique essentiel, touchant à la "chose publique" mondiale. »

Dans le discours de Jean-Paul II nous retrouvons plusieurs catégories de biens :

a) du domaine des valeurs (*paix, amour, etc.*). La valeur de la paix est référée par l'énoncé définitoire indirect, dit copulatif (*cf.* M. Riegel, cité par J. Rebeyrolle, 2000 :127) :

[16] La paix est un bien à promouvoir par le bien.

La structure sémantico-syntaxique marque une relation d'inclusion entre les N *paix* et *bien* où *paix* est hyponyme de *bien*. La même relation sémantique est exprimée dans la construction appositive *le bien de la paix*. Le domaine des valeurs est déterminé en discours par l'adjectif *moral* :

[17] « Le bien moral naît de l'amour, se manifeste comme amour. »

b) les choses appartenant à la nature : *les biens de la terre, les biens créés*. La relation d'équivalence entre ces deux SN est établie en discours par le lexème *usage* :

[18] « [...] il est nécessaire de tenir compte des implications éthiques de l'usage des biens de la terre. Le Concile Vatican II a opportunément rappelé que "Dieu a destiné la terre et tout ce qu'elle contient à l'usage de tous les hommes et de tous les peuples, en sorte que les biens créés doivent être mis en abondance à la disposition de tous, de façon équitable, sous la conduite de la justice, dont la charité est la compagne". »

c) les choses produites par l'homme, comme dans l'exemple de l'expansion de N : *les nouveaux biens provenant de la connaissance et du progrès technique*.

d) l'organisation de la vie sociale, désignée par l'unité polylexicale synonymique *bien public*. L'expansion prend la forme d'un énoncé définitoire (repéré par la marque *il s'agit de*) et les référents sont signalés par la marque *tels que* :

[19] « [...] des biens que l'on reconnaît communément comme des biens publics. Il s'agit des biens dont jouissent automatiquement tous les citoyens, sans avoir opéré de choix précis en la matière. C'est par exemple le cas, au niveau national, pour des biens

tels que le système judiciaire, le système de défense, le réseau autoroutier ou ferroviaire. »

La référence de l'expression *bien commun* dans le discours de JPII englobe donc plusieurs catégories de référents et se situe sur une hiérarchie axiologique entre la catégorie *valeur* et la catégorie *choses* (créées par Dieu et par les hommes). Nous pouvons le constater dans le passage suivant :

[20] « Pour promouvoir la paix, [...] il faut s'attacher avec une particulière attention au bien commun et à ses manifestations sociales et politiques. En effet, lorsque, à tous les niveaux, on cultive le bien commun, on cultive la paix. »

L'expression *bien commun* apparaît dans ce contexte deux fois. Dans le premier cas, *manifestations sociales et politiques [du bien commun]*, reprise anaphoriquement par le possessif *ses*, elle désigne la catégorie *ensemble de conditions de vie sociale*. Dans le deuxième cas, lorsqu'elle est en position d'argument du prédicat *cultiver*, sa référence est explicitement distinguée de la *paix*, en position d'argument du même prédicat. La catégorie *choses créées* est reliée à la catégorie *organisation de la vie sociale* par la composante *système judiciaire* repérée dans la prédication sur les *biens créés* (cf. [18] qui doivent être mis en abondance sous la conduite de la justice) et dans l'énumération des référents du *bien public*, cf. [19].

2.2.

Passons maintenant à l'étude des adjectifs qui sont en rapport de sens avec l'adjectif *commun*. L'analyse de l'adjectif *commun* et de ses synonymes nous conduit à nous interroger sur la quantité des acteurs (*créateurs, propriétaires, bénéficiaires, administrateurs*) impliqués par les prédications sur les référents qualifiés par une valeur positive. Dans les syntagmes étudiés ci-dessus nous avons constaté la présence de deux adjectifs synonymes : *public* et *collectif*.

Pour PQ, nous l'avons déjà vu, le *bien commun* est synonyme de *res publica*, ([cf. 1], « chose » *publique*) paraphrasé ailleurs *bien appartenant au domaine public*. L'adjectif *public* implique une opposition au *privé*. Cette opposition est maintenue dans une autre expression *intérêt public/intérêt privé*, paraphrasée *chose qui intéresse le public ou une personne* :

[21] « L'intérêt public est beaucoup plus difficile à définir que l'intérêt privé. C'est un concept plus abstrait. Il intéresse tout le monde, et donc personne en particulier. Plus les problèmes sont abstraits et globaux, plus ils sont difficiles à traiter et à assimiler par le public. »

Dans le discours de JPII nous avons la même expression synonymique *bien public* (cf. [19]). L'adjectif *public* n'est pas toutefois mis en opposition avec *privé*. Il n'apparaît pas non plus dans la collocation avec le N *intérêt*. Le N *intérêt* est déterminé par l'adjectif *commun*, paraphrasé comme *chose qui intéresse tous en commun* :

[22] « Dans le monde, totalement pris aujourd'hui par le phénomène de la mondialisation, les biens publics sont toujours plus nombreux à revêtir un caractère global et, par conséquent, ils augmentent aussi de jour en jour les intérêts communs. »

L'adjectif *collectif* est employé seulement par PQ au sens plus large que *commun* et *public* :

[23] « les biens collectifs mondiaux (*res communis, res nullius et res publica*).

Il détermine également le N *intérêt* :

[24] La notion de « bien commun mondial » vise l'intérêt collectif de tous les habitants du monde.

Quel sens de l'adjectif *commun* se dégage-t-il par opposition à ses synonymes ? Dans le discours de PQ il est paraphrasé dans le sens quantitatif, conformément à l'énoncé définitoire [1] et introduit la composante *propriétaires (tous)*. Rapproché de l'adjectif *public*, c'est l'opposition *tous/individu* qui est mise en relief ; rapproché de *collectif*, c'est l'extension de *tous* qui est signalée et explicitée. Le sens de l'adjectif *commun* est différent dans le discours de JPII. Le sens quantitatif, actualisé davantage par l'adjectif *public*, est secondaire par rapport au sens paraphrasé par un adverbe *communément* ou par la locution prépositionnelle *en commun*, notamment avec les prédicats cités en [4] : *cultiver, réaliser, (s'engager) communément (pour) le bien qui appartient à tous*.

Les adjectifs qui constituent le contexte immédiat du N référant à des *choses qualifiées positivement* ouvrent pour ainsi dire une possibilité d'argumenter sur le nombre d'acteurs et sur leur façon d'agir par rapport à un *bien qui les concerne*.

3. VERS DE NOUVELLES PERSPECTIVES D'ÉTUDE DU SENS DE L'EXPRESSION *BIEN COMMUN*

Les analyses de l'unité polylexicale *bien commun* sur l'axe syntagmatique et paradigmatique nous permettent de constater que nous avons affaire à deux unités linguistiques d'ordre différent : 1. *bien commun* qui est une collocation dans le discours de PQ car cette unité n'est pas figée, le N est un hyperonyme et désigne une catégorie de référents générale, et il est possible de remplacer l'adjectif collocatif par un autre sans modifier la référence de la base, même si elle est exprimée par le N métonymique *bien* et sa traduction littérale *res* ou *chose* ; 2. *bien commun* qui est un nom composé exocentrique dans le discours de JPII car il n'admet aucune variante, le N n'est pas un hyperonyme et désigne une réalité sociale dynamique dans sa complexité entre les choses et les valeurs, et il n'est pas possible de remplacer l'adjectif *commun* par l'adjectif *public* sans entraîner le changement de la référence, la structure morphologique synthétise une paraphrase prédicative plus complexe : *ensemble de conditions de la vie sociale concernant un bien auquel s'engagent communément toutes les personnes*.

Nous avons également vu que les prédicats verbaux associés au *bien commun* et extraits des deux discours [3-9] diffèrent non seulement par le choix lexical mais aussi par la place du *bien commun*, ce qui signale que la construction du sens se fait à un autre niveau que le niveau linguistique formalisé selon les principes distributionnels. Le sens de ces unités se

construit d'après une proximité qui ne peut pas se réduire à un simple assemblage de concepts associés aux lexèmes en contexte linguistique immédiat. L'unité polylexicale *bien commun*, peu importe si elle a un statut linguistique de collocation ou de nom composé, l'un ou l'autre devenant un terme en discours spécialisé, quand elle est analysée en discours, est une « formule » selon les termes de P. Fiala (2002) car employée fréquemment dans un espace public, elle est l'objet de connaissances partagées et à ce titre elle renvoie (sur le plan linguistique) aux questions de catégorisation nominale, de construction référentielle, de pragmatique lexicale et d'argumentation.

La méthodologie qui semble être la plus appropriée à la description de la construction du sens d'une formule est celle de la « sémantique référentielle ouverte sur la cognition » de G. Kleiber (1994, 1999) et celle qui adopte la conception cognitive de profil et de perspective en termes référentiels (R. Grzegorzczkowska 2004, D. Śliwa 2004a, b). Selon cette méthodologie, nous pouvons déduire de l'analyse morphologique, telle que nous venons de la présenter, une représentation conceptuelle des référents impliqués par les relations (souvent d'ordre argumentatif) établies en un discours donné. Ainsi, nous pouvons analyser la prédication (attribution d'une propriété à une entité du réel) sur un référent à partir d'un acteur. Dans le discours de PQ, les prédicats verbaux (cf. [4]) indiquent l'activité d'un acteur qui peut être un créateur en fonction du référent *chose* susceptible d'être le résultat de l'action désignée par le prédicat *créer*, un administrateur (*gérer un bien commun/public/collectif*) ou tout autre bénéficiaire (*respecter*). La prédication sur le référent *ensemble de conditions de la vie sociale concernant un bien auquel s'engagent communément toutes les personnes* dans le discours de JPII est plus complexe. À partir de la perspective d'un acteur, les prédicats *cultiver*, *exiger*, *s'engager pour* vont tous dans le sens d'une participation active des acteurs (aussi bien *créateurs* que *bénéficiaires*). Dans cette perspective, nous notons aussi une prédication (*Les visions de la réalité humaine transforment le bien commun en un simple bien-être socio-économique et le dépouillent de sa plus profonde raison d'être*) par la métonymie *personne – attribut* sur les lexèmes désignant l'acteur (il s'agit en effet des personnes qui ont une certaine vision de la réalité). La prédication à partir du référent est une attribution d'une propriété au référent, et plus exactement à un profil donné du référent. Ainsi, les prédications analysées dans les exemples [6-9] se répartissent en deux profils : a) celui de la relation aux acteurs impliqués dans l'organisation – prédiquée par le verbe au sens propre (*le bien commun concerne une personne*) ou par le verbe au sens métaphorique (*le bien commun exige le respect et la promotion de la personne*), b) celui de la dimension du référent – prédiquée par un verbe support (*prendre une dimension*, cf. [9]) ou par un verbe métaphorique (*revêtir une dimension* cf. [8]).

Nous sommes consciente que la réflexion sur l'approche référentielle est à peine esquissée et qu'elle constitue un point de départ pour une étude plus systématique qui pourrait vérifier notre intuition. Nous espérons toutefois avoir convaincu que certaines unités lexicales, telle le *bien commun*, nécessitent une définition au sein du discours dans lequel elles sont employées pour déterminer leur référenciation et pour décrire la construction de leur sens.

BIBLIOGRAPHIE

- Clas (A.) et Gross (G.), 1997 : « Les classes d'objets et la désambiguïsation des synonymes » dans *Cah. Lexicol.*, 70, pp. 27-40.
- Fiala (P.), 2002 : « Formule » dans Chauradeau (P.), Maingueneau (D.), édit. : *Dictionnaire d'analyse du discours*, Paris, Éditions du Seuil, pp. 274-275.
- Grzegorzczkova (R.), 2004 : « Punkt widzenia nadawcy w znaczeniach leksemów » dans Bartmiński J. et al. (red), *Punkt widzenia w języku i w kulturze*, Lublin, Wydawnictwo UMCS, pp. 161-176.
- Kleiber (G.), 1994 : *Nominales. Essais de sémantique référentielle*, Paris, Armand Colin.
- Kleiber (G.), 1999 : *Problèmes de sémantiques. La polysémie en questions*, Lille, Presses universitaires de Septentrion.
- Rebeyrolles (J.), 2000 : *Forme et fonction de la définition en discours*, doctorat nouveau régime, Université Toulouse-Le-Mirail.
- Śliwa (D.), 2004a : « Parlons de l'espérance. Les relations sémantiques entre des verbes collocatifs en polonais et en français » dans Bogacki, K., Giermak-Zielińska-T. (édit.), *La linguistique romane en Pologne : millésime 2004*, Łask, Oficyna Wydawnicza Leksem, pp. 263-272.
- Śliwa (D.), 2004b : « Les points, les perspectives et les profils en linguistique », *Actes du 6^e Séminaire international d'études doctorales, Brno, 5-8 février 2003, Masarykova Univerzita v Brne*, pp. 189-193.

LE STÉRÉOTYPE, DU MOT AU CONCEPT : SAISIE À TRAVERS DES CONTEXTES

Said Mosbah
RLM, ISET'COM
Tunis, Tunisie
said_m_tn@yahoo.fr

Résumé : *D'abord terme technique désignant un procédé typographique, « le stéréotype » est devenu un mot d'usage courant mentionné dans les dictionnaires surtout sous une acception figurée. Puis, utilisé au cours du XX^e siècle en sociologie et en psychologie, il est vite emprunté par les critiques littéraires et tend à s'imposer comme une notion linguistique.*

Derrière la variété de ces usages, la disparité des faits qu'il désigne et le flou qui accompagne son émergence en tant que concept, il y a lieu de s'interroger sur les invariants sémantiques de son contenu qui auraient orienté son passage d'un usage à un autre et sur les aspects innovants ou spécifiques qui ont servi à le consacrer dans chaque domaine d'étude particulier.

Les différentes définitions du stéréotype, bien qu'elles forment des contextes minimaux d'emploi nécessaires à la délimitation de sa variation, ne suffisent pas pour déterminer sa portée réelle particulièrement quand il s'agit de faits de langue.

Nous nous proposons donc d'opérer des saisies dans des contextes où sont considérés des énoncés illustratifs afin de tracer le parcours de cette évolution.

Mots-clés : *terme, stéréotype, contexte, concept, mot*

1. INTRODUCTION

D'abord terme technique désignant un procédé typographique, « le stéréotype » est devenu un mot d'usage courant mentionné dans les dictionnaires. Utilisé au cours du XX^e siècle en sociologie et en psychologie sociale, il est vite emprunté par les critiques littéraires et tend à s'imposer comme une notion linguistique.

L'hétérogénéité de l'emploi et la diversité des domaines le rendent problématique à plus d'un titre : d'abord, il y a le vague de son contenu qui, selon Charles Grivel (1998 : 58), rend inutile de « rechercher [sa] définition passe-partout ». Il faut, d'ailleurs, pour qu'une telle définition soit possible, distinguer son acception générique¹ de ses acceptions particulières. Rien qu'à voir les sèmes génériques utilisés pour le définir en rapport avec les faits de langue, on se rend compte de l'ampleur d'un tel vague,

¹ « formes de stéréotypie : clichés verbaux, stéréotypes comme schèmes collectifs figés, idées reçues » (Ruth Amossy 1998 : 25)

- il est « représentation simplifiée »/« schème collectif figé »/« modèle culturel »/« image toute faite » chez Amossy (1998 : 24) ;
- « idée conventionnelle » chez Putnam (1975) ;
- « forme d'impensé »/« syntagmes figés »/« opinion banale »/« un monstre » chez Barthes (1975 : 164 ; 1965 : 92) ;
- « parole de pouvoir »/« phrases mécanisées »/« quelque chose que l'on repère »/« quelque chose à quoi tout le monde consent et tout le monde consomme » chez Pierre Barbéris ;
- « suite ouverte d'énoncés » chez Bernard Fradin ;
- « suite ouverte de phrases » chez Jean-Claude Anscombe etc.

Ensuite, il y a la profusion des termes qui entretiennent avec lui des rapports d'affinité, d'inclusion ou même parfois de synonymie, à ne citer que *cliché*, *poncif*, *lieu commun*, *idée reçue*, *langue de bois*, *prototype*, *formules banales* etc.

Enfin, il y a le désaccord quant à la dépréciation du terme. Isabelle Rieusset n'y voit qu'« un facteur à la fois de banalisation et d'illusion sur la nature même du langage » (1994 : 22) et le qualifie même de « crime contre l'humanité ». Denis Slakta (1994 : 45), s'interrogeant sur l'existence de bons stéréotypes, trouve la réponse sous la plume de Voltaire : « ce sont ceux que le jugement ratifie quand on raisonne ». Ruth Amossy et Anne Herschberg Pierrot constatent, par contre, que la notion de stéréotype devient dans certaines théories, surtout en psychosociologie, « l'objet d'une réhabilitation qui permet de souligner ses fonctions constructives ». Les définitions qui y sont proposées, tendent à « relativiser ou neutraliser les aspects dévalorisants du phénomène » (1997 : 28). D'où la bivalence constitutive de la notion de stéréotype dans la pensée contemporaine.

N'ayant l'intention ni de définir le mot, ni de reconstituer le parcours de son évolution, nous nous proposons d'abord de voir, au-delà de l'éclatement de cette notion, s'il y a un ou des invariant(s) sémantique(s) qu'aurait conservé le mot à travers la multitude de ses emplois et des contextes dans lesquels il apparaît, tout en délimitant les traits nouveaux ou spécifiques qui ont servi à le consacrer dans chaque domaine. Puis, nous analyserons, dans un même domaine (la linguistique), l'émergence d'une nouvelle acception de la notion à partir d'une autre à laquelle elle s'apparente. L'objectif est de montrer comment le cadre théorique constitue un contexte terminologique, au sens large, où les contenus des termes s'imbriquent et s'impliquent mutuellement.

Soulignons d'abord qu'à la suite de Fabienne Cusin-Berche (2003 : 20), nous utilisons le terme « contexte » pour renvoyer « à tous les éléments cognitifs, situationnels ou intertextuels susceptibles d'intervenir dans le processus de construction du sens ».

2. LE STÉRÉOTYPE : DU MOT AU CONCEPT

Si l'on consulte les dictionnaires, il paraît que c'est l'adjectif « stéréotypé » qui s'apparente le plus à l'origine typographique du nom :

- « qui paraît sortir d'un moule, tout fait, figé. *Les sottises stéréotypées...* » (*Le Petit Robert*) ;
- « qui se présente toujours sous la même forme, figé. *Gestes stéréotypés* » (*Le Petit Larousse*). Dans ces définitions, on constate la présence de trois traits caractérisant également la signification technique du terme originel ; Il s'agit du « préconstruit », de la « répétition » et de la « fixité », tous présents dans la définition suivante du stéréotype, procédé typographique :
- « Cliché typographique obtenu par coulage qui permet la reproduction de formes imprimantes au moyen de flancs » (*Le Petit Larousse*).

La nouvelle acception est étroitement liée à la symbolique de ce mode de production littéraire qui devient norme culturelle. Ainsi, selon les termes de Ruth Amossy (1991 : 26), « de la standardisation industrielle, [le stéréotype] passe à la mécanisation de la production culturelle ».

Entre les deux dictionnaires cités, les définitions relatives au nom présentent déjà une certaine différence :

- « Opinion toute faite, réduisant les singularités » (*Le Petit Robert*) ;
- « formules banales, opinion dépourvue d'originalité » (*Le Petit Larousse*).

D'abord, il y a apparition de deux sèmes nouveaux : le premier est relatif à « l'association » entre deux objets ou concepts puisque toute opinion est à propos de quelque chose ; le second est apparent dans l'explicitation plus ou moins accentuée du trait « péjoratif » (respectivement « réduisant les singularités » et « banales »/ « dépourvue d'originalité »). Puis, il y a la variation du domaine référentiel : à côté de celui des idées, *Le Petit Larousse* ajoute celui de la langue (formules).

Dans l'ensemble, nous pouvons affirmer que parmi les trois traits cités plus haut, l'introduction du mot dans le langage commun est surtout marquée par l'apparition du sème de la péjoration qui, selon Isabelle Rieussset-Lemarié (1994 : 25), est « d'autant plus forte qu'elle a fini par jouer le rôle non plus seulement d'une connotation négative mais de [sa] dénotation pure et simple ».

Cependant, l'hétérogénéité de son emploi et la diversité des domaines qu'il implique, imposent de dépasser de telles assertions puisque « tout changement du cadre référentiel exerce une influence sur l'appréhension du sémantisme de l'unité lexicale concernée par cette transposition et devient susceptible de favoriser la mise à jour des sèmes spécifiques occultés précédemment » (Cusin-Berche 2003 : 21).

Étant « à la croisée des sciences sociales et de la littérature » (Amossy), la notion de stéréotype devrait, donc, à travers ses multiples acceptions, garder au moins un invariant sur lequel se grefferaient les nouveaux sèmes porteurs de nouvelles significations. C'est de ces traces de stabilisation et de changement que nous lançons la quête.

À commencer par le domaine de la critique littéraire, nous distinguons une tendance nette à envisager le stéréotype sous l'angle de la péjoration, en conformité avec l'usage courant du mot.

Les invariants autour desquels gravite cette connotation sont presque toujours le « préconstruit » et « la fixité » qui deviennent synonymes de « banalité » ou plus « vulgarité » et « bêtise » chez Barthes (cité par Amossy 1997 : 63). Que le stéréotype soit appréhendé en rapport avec la pensée ou le langage², il y a récurrence des mêmes traits péjoratifs bien que les sites soient parfois différents (le discours/la langue).

Un tour d'horizon des définitions qui en sont données confirme cette constante. Charles Grivel constate que, chez les « modernes », *cliché* et *stéréotype* sont entendus en rapport avec « ce que la mécanique, la reprise infinie, l'anonymité, la banalité etc. empreint » (1998 : 59) ; Jean Moliné (1998 : 35) apparente le cliché par sa banalité, sa redite et le préjugé qu'il véhicule au stéréotype.

Quant à « la répétition », elle est plutôt liée au figement des expressions : Pierre Barbéris évoque comment « le contre-stéréotype devient lui-même stéréotype : « Oh, là, là, Thierry, que d'émotion ! » et surtout « Tout à fait, Thierry » phrases d'abord inventées puis mécanisées, écrit-il, sont contre-utilisées sur le mode parodique mais le procédé devient stéréotypique – Morale » (1994 : 11). Amossy fait de la répétition un trait distinctif entre le cliché et le stéréotype « qui ne procède pas par répétition littérale » (1998 : 24) étant donné qu'il s'ancre essentiellement dans le discours. Le stéréotype apparaît donc comme une construction de lecture où l'image de la femme ou tout autre thème stéréotypé est reconstituée à partir d'un processus de déchiffrement complexe.

Dans une perspective de littérature comparée, D.-H. Pageaux (cité par Amossy et Herschberg Pierrot 1997 : 70) distingue « l'image à proprement parler, qu'il définit comme la représentation d'une réalité, du stéréotype, qu'il considère [...] comme une image réductrice, monosémique (elle transmet un message unique), essentialiste (les attributs reflètent une essence du groupe) et discriminatoire (elle est liée au préjugé et au refus de la différence) ». Cette conception rejoint en partie celle de Grivel en ce qu'elle introduit deux nouveaux sèmes ; l'un, explicite : il s'agit de l'écart par rapport aux propriétés présupposées du référent puisque « réductrices » et l'autre, implicite puisque le caractère réducteur présuppose une catégorisation. En même temps, elle projette la notion en dehors des domaines de la pensée et du langage : le préjugé et la discrimination relèvent du domaine des attitudes et du comportement.

Cet élargissement pose, d'emblée, la problématique du stéréotype dans les domaines de la sociologie et de la psychologie sociale.

Dans ces deux domaines, l'étude du phénomène peut désormais être d'ordre empirique. Dans *Le Thesaurus MeSH*, le stéréotype est défini comme « une croyance concernant les caractéristiques, attributs ou comportements d'autrui comme membre de certains groupes auxquels nous n'appartenons pas ».

² Francis Grossmann, (cours de sémantique, www.u-grenoble.fr/grossmann/Cours/Chap1.htm, 2003) récusant la distinction faite par Charlotte Schapira (1999 : 2) entre « stéréotypes de pensées » et « stéréotypes de langue » précise que « la pensée » s'incarne toujours d'une manière ou d'une autre dans la langue » (page 7) et donc, il conviendrait mieux de situer la distinction « entre les stéréotypes qui s'incarnent ou non dans des expressions plus ou moins figées ».

Ici, tous les sèmes péjoratifs sont, provisoirement, suspendus au profit d'une vision neutre où il est question d'« association » et plus implicitement de « préconstruit ». En fait, cette neutralisation est due à deux facteurs.

D'abord, il y a la distinction, fondée empiriquement³, entre les stéréotypes qui sont plutôt « des reproductions mentales de la réalité » (Plous 2003), les préjugés qui sont définis comme impliquant « un jugement *a priori*, souvent négatif d'un groupe ou de ses membres » (Fiske 1998 ; Jones 1997 et Nelson 2002) et la discrimination impliquant « que les membres d'un groupe sont désavantagés ou qu'ils sont traités injustement à cause de leur appartenance à ce groupe » (Plous 2003).

Puis, il y a l'intérêt accordé aux mécanismes cognitifs et culturels qui sous-tendent l'émergence du stéréotype. Selon l'approche psychologique, il s'agit d'un processus naturel qui « résulte de notre incapacité à retenir et à interpréter toutes les données de notre environnement [...]. Le stéréotype permet donc de s'orienter et de s'adapter face à la masse d'informations qui nous submerge à chaque instant en les organisant par catégorie opérationnelle » (*Thesaurus MeSH*)⁴.

Cette dimension cognitive rattache deux sèmes nouveaux à la notion de stéréotype ainsi appréhendée : « la sélection » et « la généralisation » qui constituent l'essence du processus de catégorisation. La sélection provient du mécanisme de simplification qui fait que, face à une large masse d'informations, nous avons tendance à ne retenir que les données saillantes.

Ce mécanisme général « peut mener à des distorsions systématiques de notre perception et, parfois, à utiliser des préjugés et des stéréotypes » (Plous 2003). La généralisation est liée aux raccourcis mentaux qui, dans le cas des stéréotypes, associent aux membres d'un groupe, des images « parfois positives » et « la plupart des temps, négatives et persistantes ». Toutefois, notons que les traits de « fixité » et de « conventionnalité », bien que, souvent mentionnés dans ces études, sont néanmoins relativisés dans une visée prescriptive.

Les traits péjoratifs, occultés de la définition précédente, sont récupérés mais relativisés dans l'étude des fonctions psychologiques et sociales du stéréotype. Par exemple, on peut lire dans *Le Thesaurus MeSH* qu'il peut par exemple servir à :

- « organiser l'environnement ;
- protéger une conception du monde (exemple : défense de notre idéologie individualiste en décrivant un groupe social comme grégaire et indifférencié) ;

³ « Lorsqu'un groupe ethnique est stéréotypé à l'aide d'un attribut neutre ou positif tel qu'« orienter vers sa famille », préjugés et discrimination peuvent ne pas être impliqués. De même, un préjugé général contre les « étrangers » ou les « amputés » peut ne pas inclure de stéréotypes particuliers ou d'actes de discrimination. Parfois, il est possible qu'il ait de la discrimination de façon intentionnelle ou non, sans pourtant qu'il y ait de préjugés ou de stéréotypes » (Plous 2003).

⁴ « Exemple : Se retrouvant dans un lieu public pour un rendez-vous à une heure précise et constatant qu'elle est sans montre, une personne choisit de s'adresser, de préférence à une vieille dame qui lui semble avenante plutôt qu'un jeune skinhead qui lui paraît agressif » (*ibidem*).

- expliquer subjectivement ou arbitrairement la complexité du réel (exemple : les immigrés sont responsables du chômage) etc. ».

Première conclusion, la notion de stéréotype est sensiblement empreinte par le cadre spécifique de chaque domaine d'étude.

Dans les critiques littéraires, la notion s'apparente plus au sens commun en mettant en exergue les traits péjoratifs du stéréotype qui est surtout saisi comme effet de style, de culture ou d'époque.

Dans les sciences sociales, d'une part, le recours à l'analyse des processus cognitifs qui sous-tendent l'émergence du stéréotype et, d'autre part, l'adoption d'une démarche empirique, ont fini par dissocier le fait en lui-même de ses effets possibles, permettant ainsi de lui donner un contenu notionnel plus précis.

Toutefois, dans le même domaine, la signification du terme n'est pas exempte de variation, même dans des théories dont l'une se revendique de l'autre. L'appréhension du stéréotype chez Putnam et Anscombe nous servira d'illustration.

3. LA VARIATION NOTIONNELLE DU STÉRÉOTYPE

J. C. Anscombe inscrit sa théorie sur les stéréotypes dans la lignée de Hilary Putnam. En plus du « stéréotype », il se sert d'une autre notion empruntée, écrit-il (2001 : 60), entre autres à celui-ci : celle de « communauté linguistique ». En nous tenant aux contenus notionnels donnés à ces deux termes, dans les deux théories, nous comptons montrer comment la notion de « communauté linguistique » sert de cadre conceptuel où se déploie le contenu notionnel spécifique du stéréotype et comment toute variation au niveau de l'appréhension de cette communauté est corrélée à une variation des propriétés attribuées au stéréotype dans chaque théorie.

Pour Putnam, la signification d'un mot est un vecteur de composants dont le stéréotype n'est qu'un élément (à côté du marqueur syntaxique, du marqueur sémantique et de l'extension). Les traits rattachés à cette signification sont tous présents à l'échelle de la communauté linguistique dans sa globalité (note 8, page 11)⁵.

La notion de « communauté linguistique » chez Putnam est indissociable de la division du travail linguistique entre experts et profanes. Elle implique directement la distribution du savoir constituant la signification d'un terme entre les différents locuteurs selon leurs rôles dans une telle division. L'hypothèse fondatrice est que « chaque communauté linguistique exemplifie le type de division du travail linguistique décrit, c'est-à-dire, possède au moins quelques termes dont les « critères » associés sont connus par un sous-ensemble des locuteurs qui acquièrent les termes, et dont l'usage par les autres locuteurs dépend d'une

⁵ « *The features that are generally thought to be present in connection with a general name necessary and sufficient conditions for membership in the extension, ways of recognizing if something is in the extension ("criteria") etc. – are all present in the linguistic community considered as a collective body; but that collective body divides the « labor » of knowing and employing these various parts of the « meaning » [...] » (Putnam 1975 : 145).*

coopération structurée entre eux et les locuteurs dans les sous-ensembles pertinents » (1975 : 146).

Le premier résultat immédiat de cette coopération se manifeste à travers la conventionalité de l'usage et la stabilité relative des composants de la signification y compris le stéréotype (la variation n'est sentie qu'à travers le temps et les époques). Selon le type de composant concerné, cette conventionalité peut être décrite à deux niveaux différents.

L'extension fait l'objet d'une double conventionalité : la première est sociolectale et explicite, puisque le savoir qui lui est relatif n'est partagé que par les experts ; la seconde est implicite, puisqu'elle n'est extensible à l'ensemble de la communauté que dans le cadre de la division du travail linguistique.

Les autres « composants du vecteur représentent tous une hypothèse au sujet de la compétence du locuteur individuel » (1975 : 191). Donc, le savoir qui leur est relatif est partagé de tous à chacun – profanes et experts. C'est dans ce cadre que le stéréotype, défini par ailleurs comme « une idée conventionnelle associée à un mot », s'oppose à l'extension en termes d'écart entre les représentations partagées plus ou moins vraies et les propriétés du référent issues d'une expertise scientifique.

Le second résultat émanant de cette coopération consiste dans le fait que la communauté linguistique, unifiée et dans son ensemble, détermine, conformément à la syntaxe et au lexique de la langue, le savoir minimum et standard mis en circulation dans la communication (1975 : 168). Au niveau de l'appréhension du stéréotype, ce savoir minimal se traduit par le nombre fini de ses traits sélectionnés parmi les principaux⁶ du savoir utilisé ou disponible ; d'où son caractère obligatoire.

Pour Anscombe, le stéréotype d'un terme est « une suite ouverte de phrases attachées à ce terme, et en définissant la *signification*. Chaque phrase du stéréotype est, pour le terme considéré, une *phrase stéréotypique* » (2001, 60).

En outre, la communauté linguistique est, pour lui, « tout ensemble de sujets parlants qui est présenté comme partageant (entre autres choses) une certaine liste de termes affectés des mêmes significations » (*ibidem*).

À première vue, cette communauté aurait un rôle dans la fixation du lexique et acquerrait au stéréotype une certaine stabilité. Mais, tel n'est pas le cas puisqu'« il ne s'agit pas [...] nécessairement de communautés linguistiques réelles, mais présentées comme telles et relatives aux circonstances d'énonciation, et donc au(x) domaine(s) dont il est question dans la conversation » (*ibidem*).

Le locuteur ne fait donc pas partie d'une seule communauté linguistique mais de plusieurs et chacune d'elles peut affecter à un terme une signification différente des autres. C'est ce qui explique d'une part, l'inconstance de la signification possédée par un locuteur donné : « il peut se faire que [celui-ci] la modifie pour une raison ou une autre » (*ibid.* : 61) et, d'autre part, toujours pour ce même locuteur, « la signification peut être constituée « de

⁶ « Not all criteria used by the linguistic community as a collective body are included in the stereotype. » (Putnam 1975 : 147).

phrases stéréotypiques... antinomiques » (*ibidem*) utilisées chacune dans des circonstances énonciatives qui lui sont appropriées.

En somme, si « le niveau stéréotypique [donc la signification] apparaît [...] comme régissant le fonctionnement de la langue en tant que pratique des locuteurs individuels » (*ibid.* : 58), cela signifie que la notion de communauté linguistique telle qu'elle est définie par Anscombe n'a aucun pouvoir réel sur la fixation de la signification ; elle est plutôt elle-même le produit d'un croisement variable et hasardeux, tributaire des contextes d'énonciation et des usages. Le contenu du stéréotype est, ainsi, caractérisé par :

- une liste ouverte de phrases ;
- un caractère variable d'un locuteur à un autre et chez le même locuteur selon les circonstances d'énonciation. Dans cette optique, si le stéréotype fait l'objet d'un consensus, ce ne serait que d'une manière contingente et relative à une situation d'énonciation bien déterminée ;
- l'absence de conventionnalité, puisque celle-ci présuppose un consensus durable et stable au sein d'une communauté linguistique au contour bien défini ; lequel consensus détermine l'usage.
- la notion d'écart n'est plus définie en rapport avec des propriétés de quelque nature qu'elles soient mais en rapport avec la volonté du locuteur et les circonstances de l'énonciation.

En définitive, le contenu du stéréotype varie selon l'appréhension qu'on a de la communauté linguistique. Si chez Putnam, celle-ci a un rôle stabilisateur en harmonie avec les traits finis, conventionnels et obligatoires du stéréotype, elle est chez Anscombe inconstante et contingente selon les contextes d'énonciation, d'où le caractère ouvert, non conventionnel et plus ou moins idiolectal du stéréotype.

4. CONCLUSION

En somme, on en conclut que les sciences sociales, en réhabilitant le stéréotype, apportent un ensemble de notions qui peut être capitalisé dans l'étude linguistique du phénomène surtout que des termes tels que « catégorisation », « sélection », « saillance », « conventionnalité » et « généralisation » sont indissociables du fonctionnement de la langue.

Cependant, l'aperçu sur l'économie de la notion en linguistique montre que, comme toute autre notion, elle est inextricable du réseau conceptuel auquel elle participe et qui, en retour, détermine une bonne part de son contenu.

Reste à étudier le concept de stéréotype, dans une théorie donnée, en rapport avec les contextes énonciatifs étroits qui illustrent à la fois son fonctionnement dans la langue et son économie dans la théorie en question.

Si la langue n'est pas nomenclature, c'est justement le fait de cette part du sens, préconstruite et relativement stable, qui se présente comme un écart par rapport à la réalité

des objets du monde auxquels elle est relative. Et, donc, elle mérite d'être étudiée loin des conclusions hâtives et des appréciations idéologisantes.

BIBLIOGRAPHIE

- Amossy (R.), 1991 : Les Idées reçues. Sémiologie du stéréotype, Nathan.
- Amossy (R.), 1998 : « Du cliché et du stéréotype. Bilan provisoire ou anatomie d'un parcours » dans *Le Cliché*, Presses universitaires du Mirail, pp. 21-28.
- Amossy (R.) et Herschberg Pierrot (A.) 1997 : *Stéréotypes et Clichés*, Nathan, 128 pp.
- Anscombe (J. C.), 2001 : « Le rôle du lexique dans la théorie des stéréotypes » dans *Langages*, n° 142, pp. 57-76.
- Barbérís (P.), 1998 : « Introduction » dans *Le Stéréotype. Crise et Transformations*, Centre de recherche sur la modernité, Université de Caen, pp. 9-13.
- Barthes (R.), 1975 : *Roland Barthes pour Roland Barthes*, Seuil, coll. « Écrivains pour toujours ».
- Barthes (R.), 1965 : *Éléments de sémiologie*, Gonthier, 1965.
- Barthes (R.), 1957 : *Mythologies*, Seuil, collection « Points ».
- Cusin-Berche (F.), 2003 : Les mots et leurs contextes, Presses Sorbonne nouvelle.
- Fiske (S. T.), 1998 : « Prejudice, stereotyping and discrimination » dans Gilbert (D. T.), Fiske (S. T.) et Lindzey (G.) (Edit.), *The handbook of social psychology* (4^e édition, pp. 357-411), New York, Mc Graw-Hill.
- Fraddans (B.), 1984 : « Hypothèses sur la forme de la représentation sémantique des noms » dans *Cahiers de lexicologie* n° 44, pp. 63-83.
- Grivel (Ch.), 1998 : « L'insu (cliché, standard et langue universelle ordinaire) » dans *Le Cliché*, Presses universitaires du Mirail, pp. 57-76.
- Jones (J. M.), 1997 : *Prejudice and racism*, (2^e édition), New York, Mc Graw-Hill.
- Moliné (J.), 1998 : « La culture du cliché : archéologie critique d'une notion problématique », dans *Le Cliché*, Presses universitaires du Mirail, pp. 35-56.
- Nelson (T. D.), 2002 : *The Psychology of prejudice*, Boston, MA, Allyn and Bacon.
- Plous (S.), 2003 : « the psychology of prejudice, stereotyping and discrimination. An overview » dans S. Plous (Ed.) *Understanding Prejudice and Discrimination* (pp. 3-48), New York, Mc Graw-Hill, traduit en français par Geneviève Quintdans et MélineCohn. www.understandingprejudice.org/apa/french/page2.htm
- Putnam (H.), 1975 : « The meaning of "meaning" » dans *Mind, Langage and Reality*, Cambridge University Press, pp. 215-271.
- Rieusset-Lemarié (I.), 1994 : « Stéréotype ou reproduction de langage sans sujet » dans *Le Stéréotype. Crise et Transformations*, Centre de recherche sur la modernité, Université de Caen, pp. 15-34.
- Slakta (D.) 1994 : « Stéréotype : sémiologie d'un concept », dans *Le Stéréotype. Crise et Transformations*, Centre de recherche sur la modernité, Université de Caen, pp. 35-45.

L'EXEMPLIFICATION BILINGUE DES *MOTS DE LA GRAMMAIRE* EN CONTEXTE CRÉOLOPHONE

Frédéric Torterat

Université Paris-Sorbonne (EA 3560)
Université d'État d'Haiti (FLA/URPP)
Paris, France/Port-au-Prince, Haiti
ft-fla@netcourrier.com

Résumé : *L'harmonisation des exigences académiques au plan grammaticographique (syntagmatique, textuel), en Haïti, ne concerne évidemment pas que les enseignements du primaire et du secondaire où le français est pratiqué généralement comme langue seconde. Aussitôt qu'il s'agit d'explicitier un phénomène linguistique et de recourir à son exemplification, les enseignants du supérieur sont confrontés à l'obligation d'en donner une représentation formelle satisfaisante. Or, dans ce contexte, la grammaire du rôle et de la référence semble présenter une appropiété descriptive incontestable.*

Mots-clés : *exemplification, FLS, créole, linguistique formelle*

1. EN TERMES DE DESCRIPTION ET D'EXEMPLIFICATION

Quand elle s'inscrit dans une démarche didactique, l'approche du fait grammatical conduit bien entendu à segmenter des phrases pour en délimiter les constituants, mais aussi des moments textuels, et donc à recourir à l'exemplification pour conforter la description d'ensemble. D'autre part, elle en appelle à des termes épilinguistiques dont l'explicitation n'est pas toujours des plus faciles, qui plus est dans le cadre d'un bilinguisme inégalement réparti, et à l'occasion duquel les supports pédagogiques ne sont entièrement rédigés dans la langue maternelle des apprenants que dans des cas limités¹. En Haïti plus particulièrement, l'exemplification des faits grammaticaux dans un enseignement bilingue (créole haïtien, français) fait l'objet d'initiatives éparpillées qui sont laissées à la discrétion des établissements et des équipes. On peut dire qu'elle sert généralement, dans les manuels du primaire et du secondaire, soit à conforter des classifiants préétablis, soit à dégager un commentaire (*kòmentè*) d'ordre général. On assiste alors à un schème d'action très répandu qui consiste à faire intervenir le champ de la grammaticographie en français, et celui des énoncés didactisés en créole haïtien (CH). Grossièrement, un même fait se répète : un

¹ À l'occasion des recherches menées depuis les années 1990 sur les créoles atlantiques de base lexicatrice romane, plusieurs formes d'«hygiène méthodologique», pour reprendre l'expression de G. Lazard, sont apparues dans la formulation plurilinguistique des exemples. Cela étant, cette problématique pose invariablement les mêmes questions : celle de la présentation d'ensemble d'une part, et celle, d'autre part, des indications notationnelles (en créole ? en français ?). Au-delà, pour une explication, l'une générale, l'autre sommaire, des créoles en tant que tels, et notamment sur les plans sociohistorique et typologique, voir resp. DeGraff 1999 et Hazaël-Massieux 2005.

domaine, celui de l'exemple, est réservé à la langue maternelle, tandis que l'autre, celui du terme grammatical, est en partie dédié à la langue seconde (FLS)². La principale difficulté pour les enseignants haïtiens consiste donc à s'approprier des termes linguistiques en français tout en les replaçant dans un contexte créolophone et, si ce n'est à les créoliser en partie, du moins à les expliciter à travers une paraphrase accessible, dont voici une illustration, relevée dans un manuel du dernier cycle du secondaire par S. Sylvestre (membre de l'URPP) dans son mémoire de master, à propos de la classe des adverbes :

[1] « *Se yon seri mo ki pa gen menm fòm, menm wòl, menm konpòtman. An teyori yon advèb diferan de yon pwepozisyon, yon pwepozisyon eswa yon adjektif menm si diferans ant mo sa yo mens pafwa, si nou pran egzanp, yon pwepozisyon konn tounen yon advèb : Jak ap vi apre. Sa pale de adjektif kalifikatif ki tounen advèb tanzantan : misye se yon nèg fò/misye pale fò³. »*

Dans un contexte pédagogique d'apprenants créolophones, et où le français intervient comme langue de la désignation taxinomique, il est courant que les chargés de cours insistent, notamment après une explication générale de la flexion verbale, sur l'unité focale (qu'il s'agisse du support agentif ou non) et, précisément, sur la portée de l'adverbe, suivant par là même en partie le système de Klein et Perdue (que résume Véronique (2003) à l'appui notamment des approches formelles : 89 *sqq* ; et surtout 97-98). Cela étant, aussitôt qu'ils le peuvent, les enseignants rebondissent sur une donnée du contexte extralinguistique pour exemplifier leur propos, et s'entendent sur des termes repris du français, qu'ils reportent en CH suivant leurs exigences pédagogiques.

Dans les manuels du troisième cycle dit « fondamental » haïtien, en effet, on déplore une terminologie flottante, y compris en créole. Par exemple pour les déterminants, tantôt nous avons *detéminan*, tantôt *mo detay* ; pour les pronoms, tantôt *pwonon*, tantôt *mo ranplasan*, avec des phrases démonstratives qui sont soustraites à toute forme de contextualisation. Le fait notamment que les adjectifs, de leur côté, soient presque tous rangés sous l'appellatif de *mo kalifyan* pose une difficulté de deux ordres : d'une part, tous les adjectifs admis dans cette catégorie ne sont pas forcément caractérisants, et, d'autre part, certains d'entre eux sont facilement nominalisables, là où d'autres font office de déterminants antéposés ou postposés, à compter que l'effectivité même de la présence d'adjectifs de la classe nominale en CH est un fait discutable (Véronique 2000). Pour pallier cette dispersion, le recours à certains phénomènes translinguistiques, comme l'intensification et la réduplication, ouvre tous les possibles. Effectivement, on note en CH une intensité nulle, notamment à l'appui de certains adverbes (*pas – menm, pa – ditou*), une intensité faible, avec des formules du type *yon tigout, yon ti jan* devant l'adjectif (*manje a ti gout sale*), et une intensité forte,

² La présentation d'ensemble fait sens en elle-même : sous une forme schématique, les distinctions formelles sont pour le moins significatives, suivant par exemple que l'on se positionne dans un cadre de réseaux relationnels ou dans celui des modules d'expression.

³ *trad.* : « C'est une classe de mots qui n'ont pas la même forme, ni les mêmes rôles (argumentaux), ni les mêmes comportements (syntagmatiques). En théorie, l'adverbe se distingue en cela de la préposition tout comme de l'adjectif, même si ces distinctions entre ces derniers se voient quelquefois contredites par les faits : ainsi une préposition devenant un adverbe (*Jack a vécu après*), voire à l'occasion un adjectif qualificatif (ce monsieur, c'est un noir fort/ce monsieur parle fort) ».

notamment avec la reduplication, laquelle se combine quelquefois avec des expressions verbales locutionnelles comme *ou kwè se* (*ou kwè se insolan li insolan*)⁴. Prenons par exemple le terme *tifi* (la fillette) : ce nom peut être visiblement adjectivé et modifié en tant que tel par un adverbe d'intensité, et ainsi en CH, dans certains contextes, on dira *tifi a tifi anpil* pour signifier que le second emploi de *tifi* reporte à l'adjectivation du premier (*tifi a*, déterminé par *a*), l'ensemble signifiant alors une fillette qui est *vraiment* fillette, autrement dit qui n'a pas encore les atours d'une femme.

Dans les ouvrages pédagogiques du secondaire et sur un plan épilinguistique en revanche, il est couramment question du *nwayo* (*yon gwoup nominal*), soit le noyau prédicateur, ou nexus, comme on voudra, ce qui représente un bon point d'appui. Dans cette vue, et afin de ne pas provoquer de rupture dans les intitulations des cours, les enseignants du supérieur emploient eux aussi tantôt *nwayo*, tantôt *pilye*, ce qui revient au même. Au moment d'aborder les syntagmes verbal et nominal, par exemple avec un noyau nominal dans *li prezidan depi ane pase*, verbal dans *li mouri yè*, nomino-adjectival dans *syel la blè* (tournure attributive averbale), ceux-ci conviennent du *wòl* (rôle argumental) que l'on pourrait résumer de modificateur des circonstants, et parlent volontiers de transitivité partout où c'est effectivement le cas (verbe, adverbe, préposition, notamment). Il en est de même pour la prédicativité, qu'ils illustrent à travers la périphrase de ce qui supporte le *wòl prensipal* dans le syntagme, et autour duquel, surtout, les modificateurs rentrent dans le cadre d'une périphérie cotextuelle donnée⁵.

C'est donc en conformité avec les démarches déjà menées dans le primaire et le secondaire, que les enseignants du supérieur, de leur côté, tranchent en faveur d'un ou de deux formalismes linguistiques parmi ceux qui leur paraissent les plus appropriés, et c'est ce point qui sera plus particulièrement traité ici.

2. CONCERNANT LE TERRAIN

Pour que les enseignants des autres degrés puissent conduire favorablement l'exemplification, en créole haïtien, de faits grammaticaux dont la désignation est dans presque tous les cas reprise du français, c'est donc au niveau de l'enseignement supérieur qu'il convient d'agir en premier temps. À ce titre, l'établissement qui accueille un programme de recherche organisé à Port-au-Prince est la faculté de linguistique appliquée. Administré par Pierre Vernet et parrainée en partie par l'AUF, il s'agit d'un établissement d'enseignement supérieur public de l'Université d'État d'Haïti. À l'occasion d'une réorganisation des cursus, un groupe de sessions 2005-2009 a été mis en place en vue, d'une part, de consolider la formation académique et pédagogique des personnels déjà en responsabilité, ainsi que de préparer, d'autre part, les recrutements à venir d'enseignants permanents et non permanents. En outre, une unité de recherche nommée Pradel Pompilus

⁴ Voir là-dessus notamment Kouwenberg 2002.

⁵ Nous avons assisté, dans les cours de deuxième année universitaire où le français s'impose encore comme langue seconde, à des explications très accessibles sur ces *advèb ki modifye mo oswa gwoup mo li akonpaye yo*. Tantôt ces derniers *modifye sans mo sa*, ou bien *sans tout fraz la*. Le terme prédicatif est alors indiqué comme recevant des *arguments* dans sa périphérie, dans un ensemble facilement formalisable.

(URPP), en hommage au linguiste haïtien renommé, a été créée en avril 2005, avec en prévision une répartition en plusieurs laboratoires à partir de 2009. Des domaines de recherche comme la créolistique (et donc la linguistique créole), la sociolinguistique, la didactique du FLE/S et la phonologie sont autant de champs disciplinaires dans le cadre desquels il est envisageable de projeter ces unités (*cf.* Torterat 2002).

Les sessions concernées ont repris les généralités de la linguistique contemporaine en termes d'approches (d'unification, notamment), et de démarches (sociohistorique, typologique), tout en exposant les champs disciplinaires à la disposition des chercheurs. Les TD, de leur côté, sont revenus sur au moins deux formes de syntactisation de l'énoncé : la parataxe, la catataxe, la diataxe et autres possibilités ont été ainsi envisagées à travers une exemplification débattue avec les chercheurs, en confrontant notamment plusieurs types de commentaires épilinguistiques et en tâchant de mesurer leur productivité (notamment d'après Eynde (Van den), Mertens et Swiggers (1998), Katzenberger et Catana-Amitay (2002)). Les enseignants concernés par le programme ont fourni par ailleurs l'inventaire de leurs intitulés de cours sous forme de syllabus avec une explication sommaire en créole haïtien, avec quelques notes pour la question du contexte.

Le principal défi méthodologique tient à ce que le fait de dégager une construction grammaticale, tout en montrant comment elle s'inscrit dans une textualité, ainsi que dans un contexte extraverbal, exige un traitement plus ou moins unifié (Torterat (2005b), (2005c)). Mettons des constructions verbales qui peuvent paraître problématiques, comme « *te konpra (n) n Potoprens se paradi* » ou « *Tobi ansanm ak Frank gen dwa gen opinyon e viv lavi* » (« tu as pris Port-au-Prince pour un paradis » / « Tobie et Frank ont droit à avoir une opinion et à vivre leur vie ») : sur quoi insister ? Comment aborder les phénomènes d'extraction, de coordination asymétrique, de réduplication ? Pour ce faire, il convient d'admettre que les supports pédagogiques ne peuvent être envisagés que dans une problématique téléonomique, et que si l'on prend une phrase pour exemplifier tel ou tel phénomène cognitif, l'un des enjeux principaux revient à la (re)contextualiser. En outre, cette projection renvoie tout simplement à une contrainte incompressible de l'enseignement : il s'agit de savoir ce qu'on entend démontrer, et pourquoi pas à l'appui de textes littéraires (Vernet 2001) ou notamment journalistiques (d'Aboville 2004), car si l'on s'en tient juste aux classifiants grammaticaux tels qu'ils sont pratiqués dans les manuels du primaire ou du secondaire, il est difficile d'échapper à la récitation et l'énumération de *règles* normatives et soustraites à toute forme de remise en contexte.

Indiquons par ailleurs que ce sont les grammaires syntagmatiques, comme on s'en doute, qui demeurent les plus sollicitées dans le supérieur haïtien, car elles placent le syntagme au premier plan. Ce type d'approche présente bien évidemment l'avantage de la simplicité, ce qui, dans le cadre d'un enseignement bilingue, permet de varier notablement les manières d'explicitier les opérations linguistiques dégagées, de même que les phénomènes cognitifs en question (ainsi la relexification ou l'ellipse de catégories fonctionnelles, comme l'indique Lumsden dans DeGraff (1999 : 129-157)). Car la problématique de l'approche méthodologique des représentations rejoint, selon toute évidence, celle des supports didactisés, à condition bien sûr qu'ils ne soient pas préconstruits en français (ce qui permet au moins d'écarter l'incongruité de certains exemples, ainsi que le sentencieux *cela se dit, mais la population ne parle pas ainsi*).

En marge des exemples et de leur contextualisation, le domaine des formalismes intervient donc au premier plan des descriptions à caractère grammatical. Dans Torterat (2005a), nous avons exposé le regain d'intérêt pour les innovations des linguistiques européennes, ainsi que pour les champs disciplinaires typiquement français, comme peuvent l'être la sémiotique des valeurs, la grammaire d'arbres polychromes, la linguistique culiolienne. Or, l'enseignement du français pâtit bien évidemment de la faible reconnaissance de ces apports généralistes outre-atlantiques, qui, même s'ils sont relayés au Canada par des chercheurs tels que Paul Laudendeau ou Diane Vincent par exemple, ne profitent pas forcément d'une presse satisfaisante de ce côté-là de l'Atlantique nord. En revanche, la plupart des chercheurs créolistes s'entendant sur quelques opérations transversales et communes à tous les créoles, comme l'analogie, la réanalyse et la grammaticalisation, nous avons là une possibilité de mise en commun non négligeable (voir respectivement Croft (2001), Haspelmath (1999) et Kriegel (2003), *inter alii*, sur ces points)⁶.

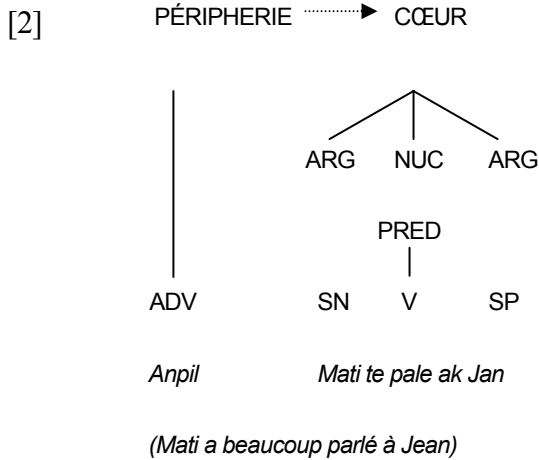
3. L'OPPORTUNITÉ MÉTHODOLOGIQUE DE LA GRAMMAIRE DU RÔLE ET DE LA RÉFÉRENCE

3.1. Présentation sommaire

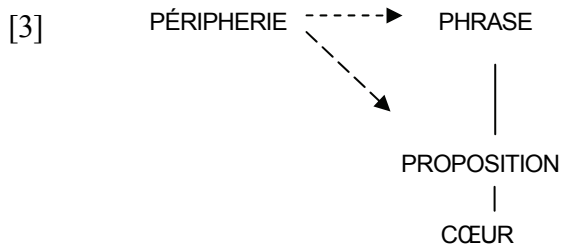
L'un des formalismes les plus à même de traiter visiblement toutes les problématiques grammaticales et de se prêter facilement à une recontextualisation des exemples didactisés, tout en les rendant accessibles, est sans doute la grammaire du rôle et de la référence. Les composantes de la GRR sont la structure L (comme en grammaire lexicale et fonctionnelle), la projection des constituants, celle des opérateurs, et enfin la pragmatique. Ces projections renvoient aux composantes descriptives des liens effectifs qui existent entre le lexique, les constituants en tant que segments syntagmatiques, les catégories (comme le temps et la quantification) et, justement, les données contextuelles (rôles topical ou focal, notamment). Dans l'énoncé, ce formalisme, que nous partions du créole haïtien ou du français, nous invite à dégager les segments syntagmatiques sur lesquels portent éventuellement des données contextuelles, à savoir un cœur (noyau de la structure argumentale) en deçà de la phrase et de la proposition, lequel cœur se répartit dans le noyau prédicatif (NUC – PRED) et ses différents arguments (ARG). Il s'agit donc d'une grammaire *distribuée* (et non dérivationnelle, comme celle préconisée dans le programme minimaliste de Chomsky), et dépourvue de mouvements, étant donné que la GRR ne prévoit aucun assemblage ni déplacement de constituants, mais donne une position principale aux constituants prédicatifs ainsi qu'à la fonction référentielle. Tout ceci est bien illustré dans les présentations à caractère généraliste de Van Valin (1999), (2001), De la Villa (2001), et François (2003b) concernant les macrorôles, Nuyts, Bolkenstein et Vet (1990), Bhat (1991), Bickel (2001), Everet (2002) pour le domaine des réseaux de relations, Klabunde et von Stutterheim (1999) pour ce qui relève de la praxis linguistique, et notamment Pino Moreno (2001), Ortigosa (2002), Wedekind, Wedekind et Musa (2002), Bellosta von Colbe et François (2003a) pour ce qui relève de la segmentation phrastique.

⁶ Nous ne discuterons pas ici des positions de McWorther sur les créoles en général, lesquelles sont suffisamment bien résumées dans Kriegel *op. cit.* (159 *sqq.*) et Hazaël-Massieux *op. cit.* (9-10).

Pourquoi inviter les enseignants du supérieur haïtien à recourir à cette « grammaire », et les inciter à en vulgariser les formalisations auprès de leurs collègues des autres degrés ? Répondons simplement que dans ce type de représentation formelle, l'approche en termes de didactisation des énoncés textuels en enseignement bilingue débouche sur une représentation d'ordre syntagmatique à laquelle est liée, par l'intermédiaire du lexique et du contexte, une représentation à la fois sémantique et pragmatique, dans un esprit de complétude et de multimodularité. L'ensemble ainsi formalisé nous conduit ainsi à aborder le « fait grammatical » dans son entier. Mettons que nous employions un modifieur comme *anpil* (*beaucoup*) en CH : quelle est sa portée ? Que modifie-t-il ? Et donc à quel niveau de la représentation va-t-on l'indiquer ? Si celui-ci porte uniquement sur le verbe, c'est par exemple au niveau du cœur que l'on va l'apposer en périphérie :



Si, dans un autre cas, c'est sur l'ensemble de l'action de prédiquer que porte le modifieur (ainsi avec les adverbes d'assertion par exemple), on fera remonter la notation au niveau du syntagme verbal dans son ensemble, voire la phrase elle-même :



Les notations de coordination et de subordination répondent aux mêmes exigences schématiques. Par exemple dans une phrase où l'on aura deux syntagmes verbaux avec un cas d'hypotaxe en périphérie de l'un des constituants, on aura la possibilité de subordonner le second syntagme à l'un des arguments du premier cœur (pour *se pa yon aza si pèp ayisien li soutni/ce n'est pas un hasard s'il a été soutenu par le peuple haïtien*) :

comme tu me déranges) en premier temps, et ensuite d'énoncer un principe, afin de décrire les éventuels mouvements qui se produisent dans l'énoncé tout en s'appuyant sur des traits comme le temps, le cas ou les rôles argumentaux, que le traitement en grammaire lexicale et fonctionnelle complète avec une série de spécifications multicatégorielles. La GRR, de son côté, rejette les formats schématiques des représentations phrastiques telles que les relations grammaticales, le module X-barre, tout en leur reprochant notamment leur inappropriété descriptive dans le domaine des invariants interlinguistiques, lesquels sont notés en GRR à travers des caractéristiques définies (Van Valin et LaPolla 1997). La conception de la construction syntagmatique, connue sous l'appellatif de structure propositionnelle plate, est basée sur le fait que cette structure se résume dans le NUCLEUS, qui contient les PRÉDICATS, et surtout le CŒUR, comme nous l'avons vu. Les notations sous-spécifiées, comme l'aspect verbal ou la négation, apparaissent donc à un niveau de nucleus du prédicat tout en renvoyant aux caractéristiques plus discursives en fin de parcours. Comme dans les autres formalismes généralisables, les indications (sous-spécifications, caractérisations) sont hiérarchisées dans ce sens où elles interviennent à des niveaux différents de la représentation syntagmatique de la phrase (on parle ainsi de *core juncture*, *clausal juncture* ou par exemple de *nuclear juncture* en anglais). Un syntagme formant phrase par exemple, tel que *Mati te pale ak Jan* [2], donnerait schématiquement PHRASE- PROPOSITION- CŒUR (ARG [SN *Mati*]-NUC (PRED) [V *pale*]-ARG [SP *ak Jan*]).

En cas d'ellipse du verbe par exemple, le mécanisme des relations renverra éventuellement à la (sous-) catégorisation commune aux constituants, et c'est dans un même esprit que les arbres représenteront autant de structures argumentales qu'il y a de SV effectifs, qu'ils soient présents ou non. Or, une exemplification de ce type dans le contexte de l'enseignement supérieur, creuset des enseignants du primaire et du secondaire, mais aussi des conseillers pédagogiques qui seront amenés à intervenir dans les établissements pour expliciter, en créole haïtien tout comme éventuellement en français, la segmentation de la phrase et l'analyse en constituants, paraît des plus appropriés.

4. EN RÉSUMÉ

Pour que les pédagogues soient en mesure, en Haïti, d'aborder avec un minimum d'harmonisation ce que nous appellerions la grammaire de phrase et la grammaire de texte, il convient qu'ils aient été habitués, en amont, à aborder ces domaines avec des représentations claires. Celles-ci leur permettront, au moment où ils dégagent les segments phrastiques (à travers les rôles argumentaux par exemple) et les moments textuels, d'exemplifier leur propos tout en apportant un volume suffisant d'indications contextuelles, et pourquoi pas à l'appui de termes qui appartiennent à un même scénario conceptuel (le *frame* de Koch). Une fois les supports recontextualisés, à la fois en CH et en FLS, ils auront la possibilité d'illustrer les opérations translinguistiques de leur choix dans une problématique d'invariabilité. Dans le supérieur haïtien, il importe donc de déterminer un ensemble fini non pas de représentations, mais d'opérations schématiques, ce qui se fait déjà en linguistique cognitive bien entendu, mais aussi en linguistique computationnelle (ainsi Stolzenburg, Höhne, Koch et Volk en appellent, dans leur programmation, à un ensemble limité d'opérations). C'est ce que nous avons dans la grammaire applicative

universelle, mais aussi, de manière plus revendiquée, dans la GRR et le programme minimaliste chomskyen. Or, dans une grammaire d'interaction (Tomasello (1998), François (2003a), Perrier (2004)) comme peut l'être la GRR, ce que l'on nommera le nexus (et/ou le *core*) du syntagme s'avère assurément significatif aux yeux des enseignants à tous les degrés d'exercice, lesquels admettent généralement la présence dans la phrase de ce qu'ils appellent un *pilier* (*pilye*), véritable donnée du contexte culturel cette fois-ci, qui se tient à l'écart des régionalismes et des phénomènes de francisation ou de pseudo-créolisation (il est question en Haïti de *pilier de famille*, *pilier de groupe*, et ce terme intègre tout simplement la grammaire de phrase de la même manière qu'il le fait de phénomènes de société). Certains enseignants, dès le secondaire, lui substituent d'ailleurs le terme de *prédicat*, ou *prédical*, pour désigner le noyau prédicateur de la phrase ou du syntagme délimité dans la phrase, et à aucun moment ne s'inscrivent en rupture avec ce formalisme, lequel dépasse la controverse, en marge du français, qui consiste à distinguer les créolophones qui seraient les tenants du CH *rèk* (unilingue et multiforme, celui des analphabètes grossièrement), et du CH *swa*, qui renverrait au parler des élèves scolarisés, avec son attirail d'alternances codiques et d'expressions littérorantes. La grammaire du rôle et de la référence présente donc une opportunité méthodologique incontestable, dans ce sens où elle s'avère pleinement compatible avec les pratiques pédagogiques du primaire et du secondaire haïtiens. Par ailleurs, celle-ci remet en cause les frontières existant entre les disciplines linguistiques, dans le même temps qu'elle prend facilement appui sur l'analyse de *corpus* textuels. Elle est également une invitation à prendre en compte le contexte d'emploi des mots tout en l'intégrant dans son appareil descriptif, au-delà de toute contrainte de normalisation des « mots de la grammaire ». Le contexte devient donc, par son intermédiaire, un pilier, là aussi, mais plus particulièrement pour ce qui relève de la construction du sens.

BIBLIOGRAPHIE

- Bellosta von Colbe (V.), François (J.), 2002 : « Windowing of attention and the core-periphery boundary » dans *Cahiers du CRISCO* 9, Université de Caen.
- Bhat (D. N. S.), 1991 : *Grammatical Relations : the Evidence against their necessity and universality*, London, Routledge.
- Bickel (B.), 2001 : « PSAs, Valence and the Theory of grammatical Relations » dans *Proceedings of the 2001 Role and Reference Grammar Workshop and Conference*.
- Croft (W.), 2001 : *Explaining language Change. An evolutionary approach*, Harlow, Longman.
- D'Aboville (F.), 2004 : *Co-construction de l'image de soi dans la presse périodique haïtienne*, mémoire de master 1, Université de Paris V-Sorbonne.
- DeGraff (M.) *édit.*, 1999 : *Language Creation and Language Change. Creolization, Diachrony and Development*, Cambridge (Mass.), MIT Press.
- Everet (D.), 2002 : « The particular vs the general Understanding of phrase structures » dans *Proceedings of the 2002 International Course and Conference on Role and Reference Grammar*.
- François (J.), 2003a : « La *Role and Reference Grammar*, une Grammaire de l'interface entre syntaxe, sémantique et pragmatique », dans *Linx*, 48, pp. 77-89.
- François (J.) *dir.*, 2003b : « Aspects de la *Role and Reference Grammar* » dans *Cahier du CRISCO*, 13, Université de Caen.

- Haspelmath (M.), 1999 : « Why is Grammaticalization irreversible ? » dans *Linguistics*, 37, 6, pp. 1043-1068.
- Hazaël-Massieux (M. C.), 2005 : « Au sujet des langues créoles » dans *La Linguistique (n° sp.)* 41, 1, pp. 3-17.
- Katzenberger (I.), Catana-Amitay (D.), 2002 : « Segmentation Marking in text production » dans *Linguistics* 40, 6, pp. 1161-1184.
- Klabunde (R.), Von Stutterheim (Ch.) (édit.), 1999 : *Representations and Processes in Language Production*, Wiesbaden, Deutscher Universitäts Verlag.
- Kouwenberg (S.) (édit.), 2002 : *Twice as meaningful. Reduplication in pidgins and creoles*, London, Batsford publications.
- Kriegel (S.) dir., 2003 : *Grammaticalisation et Réanalyse*, Paris, CNRS éditions.
- Nuyts (J.), Bolkenstein (A. M.), Vet (C.) (édit.), 1990 : *Layers and Levels of Representation in language theory : a functional View*, Amsterdam, Benjamin.
- Ortigosa (A.), 2002 : « Tense and Time Adverbials : their scope in the sentence » dans *Proceedings of the 2002 International Course and Conference on Role and Reference Grammar*.
- Perrier (G.), 2004 : « La Sémantique dans les grammaires d'interaction » dans *actes de la onzième conférence sur le traitement automatique des langues (TALN 2004)*, pp. 351-360.
- Pino Moreno (M.), 2001 : « Aplicación de Role and Reference Grammar a un estudio de variación sintáctica basado en corpus : las cláusulas pasivas e impersonales transitivas » dans *Deutscher Hispanistentag 13. Congreso de la Asociación Alemana de Hispanistas*, Universidad de Leipzig.
- Tomasello (M.) (édit.), 1998 : *The new Psychology of Language : cognitive and functional Approaches to Language Structure*, Mahwah NJ, Lawrence Erlbaum Associates.
- Torterat (F.), 2002 : *Approches de l'organisation textuelle du discours, Cahiers de la FLA*, Port-au-Prince.
- Torterat (F.), 2005a : « L'Enseignement du F.L.E. en Haïti : qu'en est-il ? » dans *Actes du colloque « Quelle Didactique de l'interculturel dans les nouveaux contextes d'enseignement-apprentissage du FLE/S ? »*, Publications de l'Université catholique de Louvain-la-Neuve (département d'études romanes), [à paraître].
- Torterat (F.), 2005b : « Le Français comme langue seconde en Haïti » dans *actes du colloque FIPF 2005 (Didactiques et convergences des langues et des cultures)*, [à paraître].
- Torterat (F.), 2005c : *Cours de linguistique modulaire, A.U.F. Ens. Sup.*, www.ht.auf.org/coursenligne/linguistique
- Van den Eynde (K.), Mertens (P.), Swiggers (P.), 1998 : « Structuration segmentale et suprasegmentale en syntaxe » dans *Analyse linguistique et Approches de l'oral, Orbis Supplementa* 10, Peeters, pp. 33-57.
- Van Valin (R. J.), 1999 : « Generalized semantic Roles and the Syntax-semantics interface » dans *Empirical Issues in formal Syntax and Semantics* 2 (F. Corblin, C. Dobrovie-Sorin, J.-M. Marandin édit., pp. 373-389.
- Van Valin (R. J.), 2001 : *Introduction to Syntax*, Cambridge, CUP.
- Van Valin (R. J.), LaPolla (R.), 1997 : *Syntax : Structure, Meaning and Function*, Cambridge, Cambridge University Press.
- Vernet (P.), 2001 : *Analyse du roman/analiz woman Eritye Vilokan, de Pierre Michel Chéry*, Port-au-Prince, FLA, collection « Anthropos ».
- Véronique (D.), 2000 : « Note sur les “ adjectifs ” dans quelques créoles français » dans *Langages* 138, Paris, Larousse, pp. 61-69.
- Véronique (D.), 2003 : « Le développement de l'expression de la négation dans les créoles français et dans l'acquisition du français » dans KRIEGEL (S.), dir., *Grammaticalisation et Réanalyse*, Paris, CNRS éditions, pp. 87-109.
- Villa (J.) de la, 2001 : « Do Adjuncts have Macro-Roles ? » dans *Proceedings of the 11th International Colloquium on Latin Linguistics*, University of Amsterdam.
- Wedekind (K.), Wedekind (C.), Musa (A.), 2002 : « RRG Universal verb Classes vs Beja verb classifications based on morphologies and textual functions » dans *Proceedings of the 2002 International Course and Conference on Role and Reference Grammar*.

L'UNITÉ PHRASÉOLOGIQUE EN DIDACTIQUE DES LANGUES ÉTRANGÈRES DE SPÉCIALITÉ : LE CAS DU DISCOURS ÉCONOMIQUE D'ENTREPRISE EN ESPAGNOL

Lieve Vangehuchten

département de communication internationale de gestion, section d'espagnol
faculté de sciences économiques appliquées, Université d'Anvers

Anvers, Belgique

lieve.vangehuchten@ua.ac.be

Résumé : À partir de l'analyse d'un corpus de discours économique d'entreprise en espagnol (120 000 occurrences) nous tentons de mieux cerner la spécificité des unités phraséologiques par rapport aux termes syntagmatiques. À ce sujet nous nous basons sur les idées et définitions formulées par la linguistique appliquée à l'acquisition d'une langue seconde/étrangère ainsi que par la terminologie. De cette façon nous élaborons une hypothèse de travail à laquelle sont soumis les trois termes les plus fréquents du corpus : empresa, producto y coste.

Mots-clés : terminologie, phraséologie, séquence syntagmatique, didactique des langues étrangères de spécialité, Languages for specific purposes (LSP), linguistique de corpus

1. INTRODUCTION

De nos jours l'étude des contextes des termes a gagné de l'importance non seulement en terminologie mais aussi en didactique des langues étrangères de spécialité. Dans l'optique où la maîtrise de la phraséologie d'une langue est symptomatique pour la compétence linguistique d'un étudiant dans cette même langue (Nation 2001, Ellis 2002, Wray 2002), la didactique et la linguistique appliquée à l'acquisition d'une langue seconde/étrangère plaident de plus en plus pour une approche phraséologique (cf. le principe de « *lexico-grammar* » de, par exemple, Schmitt & Celce-Murcia 2002, Schmitt *et al.* 2004, ou encore « *the lexical approach* » de Lewis 1993).

En ce qui concerne les langues étrangères de spécialité, cette nouvelle approche pose néanmoins le problème de comment distinguer adéquatement les unités terminologiques syntagmatiques des unités phraséologiques. Il s'avère qu'en terminologie existe la plus grande confusion quant à la dénomination et la définition de ces deux concepts. « Collocation, phrasème en français, *collocation*, *multiword term*, *compound term*, *LSP phrase*, *terminological phrase*, *phraseological term*, en anglais, *unidad fraseológica (especializada)*, *fraseologismo*, en espagnol » (Béjoint et Thoiron 1992, Thomas 1993, Cabré, Estopà et Lorente 1996, Wright 1997, Tercedor Sánchez 1999, Antia 2000, Bevilacqua 2001), ce ne sont que quelques exemples de la grande variété de noms qui désignent en terminologie les séquences lexicales composées de plus d'une forme. Cette

situation n'est guère différente en linguistique appliquée à l'acquisition des langues étrangères, où abondent également les différentes dénominations, comme le signale Wray (2002 : 9).

Dans ce travail, nous tentons de résoudre le problème énoncé à partir des résultats de l'analyse d'un *corpus* de discours économique d'entreprise en espagnol dont la taille est de l'ordre de 120 000 occurrences¹. Ce *corpus* remplit la fonction de *corpus* modélisque, c'est-à-dire que son but est de servir comme sujet d'expérience et non pas de fournir les données pour la composition d'un outil de phraséologie.

2. LES TERMES SYNTAGMATIQUES (TS)

Afin de pouvoir cerner la spécificité des TS, nous suivons, parmi les multiples approches possibles en terminologie, celle de la terminologie communicative (Cabré 1999), qui distingue entre le terme composé ou syntagmatique correspondant à un concept spécialisé défini en tant que tel et qui est donc considéré comme une unité lexicale, et l'unité phraséologique spécialisée qui est formée par plus d'une unité lexicale, dépassant ainsi le niveau lexical. Malheureusement, s'il n'est pas facile en espagnol général de décider si la lexicalisation s'est produite entièrement (par ex. Alvar Ezquerro 1993), ceci est *a fortiori* le cas des langues spécialisées, où les séquences syntagmatiques sont non seulement plus abondantes mais aussi bien plus longues. Voilà pourquoi, surtout quand la caractéristique de longueur se combine avec une fréquence extrêmement basse dans le *corpus*, il est parfois difficile de tracer une ligne nette entre les TS et les unités phraséologiques spécialisées. Comme le dit Myking (1989 : 270) : « *Compounds are not always intended for institutionalization, i. e. as a term. They may function as ad-hoc descriptive or definition-like syntactic phrases, corresponding to the text-condensing compounds of news-paper headlines, etc.* »

Malgré le fait que notre définition de ce que c'est qu'un TS n'est donc pas sans faille, nous avons quand même analysé les séquences composées dans le *corpus* suivant l'approche stipulée ci-dessus. Ainsi, nous avons constaté que les unités lexicales syntagmatiques dans le *corpus* sont majoritairement des termes, c'est-à-dire dans 82,26 % des cas². Nous avons également observé que 66,19 % de tous les termes économiques dans le *corpus* sont des TS avec un taux de couverture de 29,05 %, et 45,66 % de tous les termes auxiliaires³ avec un taux de couverture de 20,42 %, comme on peut l'observer dans le tableau suivant⁴ :

¹ Il s'agit d'un manuel de gestion des entreprises de E. Pérez Gorostegui : *Introducción a la administración de empresas* (1997).

² Ceci est d'autant plus significatif quand on prend en considération que le rapport entre le lexique général et le lexique spécialisé dans le *corpus* est respectivement de 61,02 % et 33,91 %.

³ Nous avons distingué dans le *corpus* les termes relatifs à l'économie des termes empruntés à d'autres sciences qui sont auxiliaires de l'économie, comme la statistique et les mathématiques (74,60 % de tous les termes auxiliaires), et, en moindre mesure, le droit, la sociologie, les sciences politiques, la psychologie, la biologie et la physique.

⁴ UL est l'abréviation d'« unité lexicale », OC d'« occurrence ».

	Termes économiques	Termes auxiliaires
Simple	476 UL 10.264 OC 33,81 % 70,95 %	144 UL 1.691 OC 54,34 % 79,58 %
Syntagmatiques	932 UL 4.202 OC 66,19 % 29,05 %	121 UL 434 OC 45,66 % 20,42 %
Total	1.408 UL 14.466 OC 100 % 100 %	265 UL 2.125 OC 100 % 100 %

La composition semble donc être une caractéristique essentielle du langage spécialisé en général, et de l'unité terminologique en particulier. La moyenne pondérée de la longueur des TS dans le *corpus* est de 3 éléments, avec un minimum de 2 et un maximum de 12. Cependant, la plupart se composent de 2 termes (44,75 % en moyenne), et ce sont ces termes-ci qui manifestent le taux de couverture le plus élevé (58,20 % en moyenne). Les 10 TS les plus fréquents du *corpus*, avec mention de leur fréquence absolue, sont :

[1] <i>u.m.</i>	375
[2] <i>Flujo de caja</i>	87
[3] <i>S.A.</i>	86
[4] <i>Valor actual neto</i>	66
[5] <i>Coste fijo</i>	62
[6] <i>Materia prima</i>	54
[7] <i>VAN</i>	54
[8] <i>Rentabilidad financiera</i>	47
[9] <i>Rentabilidad requerida</i>	43
[10] <i>Tipo de descuento</i>	42

Parmi les termes il y a également 504 *hapax* économiques et 102 *hapax* auxiliaires. La plupart de ces *hapax* (79,76 % et 53,92 % respectivement) sont des TS, comme l'on peut voir dans le tableau ci-dessous :

	Termes économiques	Termes auxiliaires
<i>Hapax composés</i>	402 79,76 %	55 53,92 %
<i>Hapax simples</i>	102 20,24 %	47 46,08 %
<i>Total</i>	504 100 %	102 100 %

Par conséquent, le taux d'*hapax* parmi les TS économiques et auxiliaires est très élevé (43,13 % et 45,45 % respectivement) par rapport à celui parmi les termes simples (seulement 21,43 % et 32,64 %). Cette situation se doit, en partie, au phénomène de la variabilité morphosyntaxique. Les *hapax legomena* et *dislegomena* parmi les TS manifestent plusieurs formes qui se distinguent seulement par de petites différences. Ceci peut être une autre indication, mise à part la possible confusion avec des phrases définitoires dont nous avons déjà fait mention, du fait que certains termes ne se sont pas encore figés en tant que tels⁵.

3. LES UNITÉS PHRASÉOLOGIQUES (UF)

En ce qui concerne l'identification des UF, nous nous sommes basée en première instance sur les idées et définitions formulées par la didactique et la linguistique appliquée à l'acquisition d'une langue seconde/étrangère (par exemple, Nation 2001, Wray 2002, Schmitt 2004). La bibliographie avance la fréquence comme une indication importante du caractère phraséologique d'une séquence déterminée et ce critère-ci a été amplement appliqué à l'anglais général et spécifique moyennant des techniques lexicostatistiques (Church et Hanks 1990, Biber 1993, Church *et al.*, 1994). Tout de même, il présente deux inconvénients : en premier lieu, l'application desdites techniques requiert des fréquences élevées, et donc, des *corpus* énormes⁶, mais en plus, même quand le *corpus* est considéré comme étant amplement représentatif du langage étudié, il s'est avéré que de nombreuses UF, considérées comme typiques et fréquentes, y sont souvent absentes (Kennedy 1998 : 117, Read et Nation 2004 : 32). Identifier la phraséologie d'un langage déterminé – qu'il soit général ou spécialisé – uniquement à base de la fréquence et les techniques statistiques qui en font emploi, est, par conséquent, utopique.

À part la fréquence, la linguistique appliquée à l'acquisition de langues étrangères propose d'autres caractéristiques, plutôt qualitatives, comme critères permettant l'identification tels

⁵ Au total il y a 43 cas d'alternances morphosyntaxiques parmi les termes syntagmatiques. Quelques exemples en sont (avec mention de leur fréquence absolue) : *campaña publicitaria* (1) — *campaña de publicidad* (1) ; *bienestar de la sociedad* (1) — *bienestar social* (2) (alternance d'un adjectif avec un complément nominal), *precio al contado* (1) — *precio de contado* (1) ; *en línea y staff* (2) — *de línea y staff* (1) (alternance de la préposition), *método VAN* (2) — *método del VAN* (1) (suppression de la préposition et de l'article), *ruptura de stocks* (6) — *ruptura del stock* (1) (alternance du singulier et du pluriel, et de l'absence et présence de l'article), *periodo medio de maduración económico* (2) — *periodo medio de maduración económico* (6) (alternance orthographique), *empresa cooperativa* (2) — *cooperativa* (1) (suppression du noyau), *índice de cantidades de Laspeyres* (1) — *índice de Laspeyres* (2) (suppression d'une partie de l'attribut), etc.

⁶ Afin de détecter des unités phraséologiques typiques dans un segment assez restreint du langage économique — celui du marché financier — avec des techniques statistiques, Smadja (1993) devait composer un *corpus* de 10 000 000 de mots.

que le degré d'opacité sémantique (+), d'institutionnalisation (+), de fixation (+), de fossilisation ou de pétrification (+), de complexité / compacité morphosyntaxique (+), d'exclusivité quant aux combinaisons possibles (+), et d'imprévisibilité (+). À partir de ces différents critères, nous avons analysé les séquences du *corpus* contenant les trois termes les plus fréquents : *empresa*, *producto* y *coste* (respectivement 821, 491 et 330 occurrences). Vu qu'il s'agit de trois substantifs, nous nous sommes limitée aux combinaisons de base suivantes : terme + substantif (ou à l'envers), terme + adjectif (ou à l'envers), terme + verbe (ou à l'envers) avec intercalation possible d'articles et/ou prépositions. Après avoir relevé de façon automatique les concordances⁷, l'environnement lexical jusqu'à 5 formes à gauche et 5 formes à droite de ces trois termes⁸ a été examiné de façon exhaustive et manuelle.

Il est évident que, étant donné la taille modeste du *corpus* analysé, le critère de la fréquence n'est guère utilisable. Cependant, les autres critères suggérés par la linguistique appliquée et mentionnés ci-dessus ne sont pas beaucoup plus profitables. Ainsi, le terme *empresa* forme le noyau de 200 collocations⁹ au total, mais il n'est possible que d'en identifier quelques-unes, de façon plus ou moins certaine, comme des UF :

Avec adjectif :

- *la empresa fabricante* (12), ~ *competidora* (3), ~ *especializada en* (2) pour leur fréquence ;
- *la empresa saneada y técnicamente solvente* (1) pour son opacité sémantique (le sens figuré des attributs).

Avec verbe :

- *la empresa vender* (6), *dirigir* ~ (6), ~ *competir* (4), ~ *adquirir* (3), ~ *financiar* (2), ~ *cotizar en bolsa* (2) pour leur fréquence ;
- *la empresa crecer* (9) pour sa relative opacité sémantique.

Avec substantif : quoique les collocations avec substantif soient les plus fréquentes (99 cas ou 49,5 %), il est très difficile d'en identifier les UF avec certitude. Le seul critère univoque semble être celui de la compacité morphosyntaxique : *tipo de empresa* (2) et *fusión de empresas* (1) sans article. Dans 4 cas seulement, le terme *empresa* est le noyau grammatical de la collocation. Le fait qu'il est relié à son complément par la préposition *en* – très peu fréquente en comparaison avec la dominante *de* – sans intercalation d'article, fait que nous pouvons les considérer comme des UF à coup sûr : *empresa en funcionamiento* (4), ~ *en liquidación* (3), ~ *en marcha* (1), ~ *en competencia* (1).

⁷ Moyennant le programme *Concordancer*, version 2.0, développé par l'*Institut für Sprach- und Literaturwissenschaft* de la *Technische Hochschule Darmstadt* en Allemagne.

⁸ D'après la bibliographie les combinaisons syntagmatiques typiques se donnent jusqu'à 4 ou 5 mots devant et derrière le mot-noyau (Smadja 1991 : 174 ; Stubbs 2001 : 29).

⁹ Nous utilisons « collocation » dans ce travail dans son acception la plus généralisée en terminologie (Wright 1997 : 15), c'est-à-dire comme hyperonyme des unités phraséologiques et d'autres séquences discursives.

Il est donc indispensable de mieux cerner les particularités des séquences phraséologiques spécialisées, car les caractéristiques avancées jusqu'à présent s'avèrent trop peu précises. Les études récentes en linguistique appliquée à l'acquisition de langues étrangères (Wray 2002, Schmitt 2004) en viennent toutes à la même conclusion : afin de pouvoir identifier les UF dans un *corpus*, il faudrait développer une méthode éclectique qui combine plusieurs analyses sans qu'elles soient nécessairement les mêmes dans tous les cas. Le rôle du critère de la fréquence doit être également remis en question. Wray (2002 : 30) fait à ce propos la suggestion suivante : « [...] *raw frequency is not an adequate measure of formulaicity. To capture the extent to which a word string is the preferred way of expressing a given idea [...], we need to know not only how often that form can be found in the sample, but also how often it could have occurred. In other words, we need a way to calculate the occurrences of a particular message form as a proportion of the total number of attempts to express that message.* » En d'autres mots, au lieu de se baser simplement sur la fréquence absolue, cet auteur propose d'examiner la possibilité que le même contenu sémantique ou message donne lieu à différentes réalisations.

Cette idée est intéressante et nous la retrouvons, de façon plus approfondie, en terminologie. Bien qu'en terminologie la nomenclature soit aussi diverse qu'en didactique, les critères de détection se montrent plus restrictifs, car conçus spécifiquement pour l'analyse du discours spécialisé. En terminologie, voici ce que l'on considère comme obligatoire dans une UF spécialisée (par exemple, Béjoint et Thoiron 1992, Thomas 1993, Bevilacqua 2001) :

- la présence d'un terme minimum sans que la combinaison donne lieu à une nouvelle unité lexicale ;
- l'inclusion d'un noyau qui sémantiquement peut être considéré comme événementiel ;
- la fixation morphosyntaxique ;
- une fréquence « considérable ».

À propos de ces deux derniers critères, Bevilacqua (2001), dans une étude sur les UF contenant le terme *energia* dans un *corpus* de près de 85 000 occurrences, émet qu'ils sont même trop restrictifs vu que la variation ou commutation morphosyntaxique est une caractéristique intrinsèque aussi bien des langues spécialisées que de la langue générale et que, par conséquent, la fréquence absolue dans un *corpus* de taille modeste en est négativement affectée. Voilà pourquoi elle propose de subordonner, dans une première phase, cette fixation morphosyntaxique à une structure sémantique sous-jacente afin de décompter les occurrences en liaison avec la fréquence de ladite structure. Elle réussit ainsi à identifier bon nombre de séquences en tant qu'UF alors qu'elle les avait initialement rejetées comme étant des *hapax*¹⁰.

Si nous appliquons ce principe de fixation sémantique sous-jacente en combinaison avec les autres critères terminologiques aux séquences contenant les termes *empresa*, *producto* y

¹⁰ Bevilacqua (2001 : 123) n'omet que les *hapax*, étant donnée la taille restreinte de son *corpus*.

coste dans notre *corpus*, le nombre de candidats au statut d'UF monte considérablement. Bevilacqua (2001 : 124) postule que toute UF spécialisée trouve son origine dans une structure verbale sous-jacente formée par un verbe événementiel et un complément. Ce verbe événementiel peut, par la suite, se réaliser également sous la forme d'un substantif ou adjectif déverbal. En envisageant les données présentes dans notre *corpus*, nous pensons qu'il est nécessaire d'élargir ce point de départ en permettant que le terme exerce également la fonction grammaticale de sujet et non seulement de complément. Apparemment le terme *energía* n'est présent dans le *corpus* de Bevilacqua que comme complément du verbe, mais les termes *empresa*, *producto* et *coste* expriment dans un bon nombre de cas, pas toujours grammaticalement mais bien sémantiquement, le sujet de l'action événementielle.

Nous présentons par la suite les collocations identifiées dans le *corpus* analysé comme des « candidats-UF » pour les termes *empresa*, *producto* y *coste*. Vu que la taille de notre *corpus* est un peu plus importante que celui de Bevilacqua (120 000 occurrences vs. 85 000 occurrences), nous estimons, tout comme cet auteur, que nous pouvons appliquer le critère de la fréquence de la structure sous-jacente à partir de 2 occurrences, en omettant uniquement les *hapax*. Ainsi, nous trouvons pour *empresa* les suivants résultats (classés par fréquence décroissante) :

+ verbe (comme sujet ou compl.)	+ adjectif déverbal	+ substantif déverbal	Structure sémantique sous-jacente	Fréqu. totale
<i>Producir</i> (7) <i>Crear</i> (4) <i>Elaborar</i> (4)	<i>Productora</i> (1) <i>Fabricante</i> (12)	<i>Producción de la</i> (3) <i>Productividad de la</i> (2) <i>Producto de la</i> (5)	<i>T + producir/ crear/elaborar/ fabricar</i> ¹¹	38
<i>Beneficiarse de</i> (1) <i>Obtener beneficios</i> (3) <i>Conseguir beneficios</i> (2)		<i>Beneficio de la</i> (11)	<i>T + Beneficiarse/ obtener o conseguir beneficios</i>	17
<i>Crecer</i> (9)		<i>Crecimiento de la</i> (6)	<i>T + crecer</i>	15
<i>Dirigir</i> (6) <i>Gestionar</i> (1) <i>Liderar</i> (1) <i>Regir</i> (1)		<i>Dirección de la</i> (4)	<i>Dirigir/gestionar/ liderar/regir + T</i>	13
	<i>Rentable</i> (1)	<i>Rentabilidad de la</i> (10)	<i>T + rendir</i>	11

¹¹ Nous n'acceptons pas que la variation morphosyntaxique mais aussi la commutation lexicale dans la structure sémantique sous-jacente. La synonymie existe également au niveau de la phraséologie, et ce sont surtout les caractéristiques pragmatiques de la communication spécialisée qui font qu'une forme le remporte sur d'autres (Cabré 2001).

Ofrecer (productos, servicios) (8) <i>Ofertar</i> (1)		<i>Oferta de la</i> (1)	<i>T + ofrecer/ofertar</i>	10
<i>Adquirir</i> (4)		<i>Adquisición de la</i> (3) <i>Adquirente de la</i> (2)	<i>T + adquirir</i> <i>Adquirir + T</i>	9
<i>Competir</i> (4)	<i>Competidora</i> (3) <i>Concurrente</i> (1)	<i>Competencia entre</i> (1)	<i>T + competir/</i> <i>concurrir</i>	9
<i>Decidir</i> (5)		<i>Decisión de la</i> (3)	<i>T + decidir</i>	8
<i>Diversificar</i> (3)		<i>Diversificación de</i> <i>la</i> (5)	<i>Diversificar + T</i>	8
<i>Invertir</i> (2) <i>Realizar una</i> <i>inversión</i> (2)		<i>Inversión en</i> (3)	<i>T + invertir</i> <i>Invertir en + T</i>	7
<i>Endeudarse</i> (2)		<i>Deuda con la</i> (2) <i>Endeudamiento de</i> <i>la</i> (2)	<i>T + endeudarse</i>	6
<i>Financiar</i> (4)		<i>Financiación de la</i> (2)	<i>Financiar + T</i>	6
<i>Liquidar</i> (3)		<i>En liquidación</i> (3)	<i>Liquidar + T</i>	6
<i>Organizar</i> (2)	<i>Organizativo</i> (2)	<i>Organización de la</i> (2)	<i>Organizar + T</i>	6
<i>Quebrar</i> (2)		<i>En situación de</i> <i>quiebra</i> (2)	<i>T + quebrar</i>	4
<i>Sanear</i> (1)	<i>Sano</i> (1)	<i>Salud de la</i> (1)	<i>Sanear + T</i>	3
	<i>Especializada</i> <i>en</i> (2)		<i>T + especializarse en</i>	2
	<i>Absorbente</i> (1)	<i>Absorción de la</i> (1)	<i>Absorber + T</i>	2
<i>Cotizar en bolsa</i> (2)			<i>T + cotizar en bolsa</i>	2
<i>Desarrollarse</i> (2)		<i>Desarrollo de la</i> (1)	<i>Desarrollar + T</i>	2
<i>Devenir insolvente</i> (1)	<i>Solvente</i> (1)		<i>T + devenir</i> <i>insolvente</i>	2
<i>Emitir (acciones, un</i> <i>empréstimo)</i> (2)			<i>T + emitir</i>	2
<i>Fusionar</i> (1)		<i>Fusión de las</i> (1)	<i>T + fusionar</i>	2
<i>Suspender sus</i> <i>pagos</i> (1)		<i>En situación de</i> <i>suspensión de</i> <i>pagos</i> (1)	<i>T + suspender sus</i> <i>pagos</i>	2

Pour *producto* (séquences sélectionnées parmi 124 collocations au total) :

+ verbe (comme sujet ou compl.)	+ adjectif déverbal	+ substantif déverbal	Structure sémantique sous-jacente	Fréq. totale
<i>Elaborar</i> (21) <i>Desarrollar</i> (5) <i>Crear</i> (4) <i>Fabricar</i> (4) <i>Diseñar</i> (1)		<i>Creación del</i> (7) <i>Elaboración del</i> (7) <i>Desarrollo del</i> (4) <i>Diseño del</i> (3) <i>Fabricación del</i> (2)	<i>Elaborar/ desarrollar /crear/fabricar/ diseñar + T</i>	58
<i>Diferenciar</i> (9) <i>Diversificar entre</i> (1)		<i>Diferenciación del</i> (3) <i>Diversificación del</i> (1)	<i>Diferenciar/ diversificar + T</i>	14
<i>Vender</i> (4)		<i>Venta del</i> (9)	<i>Vender + T</i>	13
<i>Introducir (en el mercado)</i> (4) <i>Implantar</i> (1) <i>Lanzar</i> (1) <i>Promocionar</i> (1)		<i>Introducción del ~ en el mercado</i> (2) <i>Lanzamiento del</i> (2) <i>Promoción del</i> (1)	<i>Introducer/ implantar/ lanzar/ promocionar + T</i>	12
<i>Generar beneficios</i> (7)			<i>T + generar beneficios</i>	7
<i>Inventariar</i> (1)		<i>Inventario del</i> (6)	<i>Inventariar + T</i>	7
<i>Adquirir</i> (2) <i>Comprar</i> (2)		<i>Adquirente del</i> (1) <i>Adquisición del</i> (1)	<i>Adquirir/ comprar + T</i>	6
<i>Distribuir</i> (3)		<i>Distribución del</i> (3)	<i>Distribuir + T</i>	6
<i>Transformar en</i> (1) <i>Modificar</i> (3)		<i>Modificación del</i> (2)	<i>Transformar/ modificar + T</i>	6
	<i>Conocido</i> (2) <i>Desconocido</i> (2)	<i>Conocimiento del</i> (1)	<i>Conocer + T</i>	5
<i>Consumir</i> (3)		<i>Consumo del</i> (2)	<i>Consumir + T</i>	5
<i>Demandar</i> (1)		<i>Demanda del</i> (3) <i>Demandante del</i> (1)	<i>Demandar + T</i>	5
<i>Identificar</i> (3)		<i>Identificación del</i> (2)	<i>Identificar + T</i>	5
<i>Ofertar</i> (1) <i>Ofrecer</i> (4)			<i>Ofrecer/ofertar + T</i>	5
<i>Aceptar</i> (2)		<i>Aceptación del</i> (2)	<i>Aceptar + T</i>	4
<i>Fijar un precio a</i> (2)		<i>Fijación del precio del</i> (2)	<i>Fijar un precio a + T</i>	4
<i>Posicionar</i> (1)		<i>Posición del</i> (3)	<i>Posicionar + T</i>	4
<i>Transportar</i> (3)		<i>Transporte del</i> (1)	<i>Transportar + T</i>	4
<i>Marcar</i> (2)		<i>Marcado del</i> (1)	<i>Marcar + T</i>	3
<i>Probar</i> (1)		<i>Prueba del</i> (2)	<i>Probar + T</i>	3

<i>Terminar</i> (2)		<i>Terminación del</i> (1)	<i>Terminar + T</i>	3
		<i>Retirada del</i> (2)	<i>Retirar + T</i>	2
		<i>Vida del</i> (2)	<i>T + Vivir</i>	2
<i>Comercializar</i> (2)			<i>Comercializar + T</i>	2
<i>Normalizar</i> (1)		<i>Normalización del</i> (1)	<i>Normalizar + T</i>	2
<i>Satisfacer</i> (2)			<i>T + satisfacer</i>	2
<i>Tipificar</i> (1)		<i>Tipificación del</i> (1)	<i>Tipificar + T</i>	2

Pour *coste* (séquences sélectionnées parmi 95 collocations au total) :

+ verbe (comme sujet ou compl.)	+ adjectif déverbal	+ substantif déverbal	Structure sémantique sous-jacente	Fréq. totale
<i>Reducir</i> (7) <i>Minimizar</i> (5)	<i>Bajo</i> (4) <i>Decreciente</i> (1)	<i>Minimización del</i> (2) <i>Reducción del</i> (2)	<i>Reducir/minimizar + T</i>	21
<i>Incurrir en</i> (11) <i>Generar</i> (2) <i>Suponer</i> (2) <i>Comportar</i> (1) <i>Correr con un</i> (1)			<i>Incurrir en/generar/suponer/comportar/correr con un + T</i>	17
<i>Calcular</i> (6)		<i>Cálculo del</i> (4)	<i>Calcular + T</i>	10
<i>Crecer</i> (1) <i>Aumentar</i> (1)	<i>Creciente</i> (1) <i>Alto</i> (1) <i>Elevado</i> (1) <i>Grandes</i> (1)	<i>Aumento del</i> (1)	<i>T + crecer/aumentar</i>	7
<i>Cubrir</i> (6)		<i>Cobertura del</i> (1)	<i>Cubrir + T</i>	7
<i>Prever</i> (2)	<i>Previsto</i> (2)	<i>Previsión</i> (2)	<i>Prever + T</i>	6
<i>Imputar</i> (2)	<i>Imputable</i> (1)	<i>Imputación del</i> (2)	<i>Imputar + T</i>	5
<i>Controlar</i> (1)		<i>Control del</i> (3)	<i>Controlar + T</i>	4
<i>Determinar</i> (3)		<i>Determinación del</i> (1)	<i>Determinar + T</i>	4
<i>Conocer</i> (3)			<i>Conocer + T</i>	3
<i>Deducir</i> (2) <i>Descontar</i> (1)			<i>Deducir/descontar + T</i>	3
<i>Recuperar</i> (3)			<i>Recuperar + T</i>	3
<i>Sumar</i> (2)		<i>Suma del</i> (1)	<i>Sumar + T</i>	3
<i>Valorar al</i> (2)		<i>Valoración al</i> (1)	<i>Valorar al + T</i>	3
		<i>Estructura de (su)</i> (2)	<i>Estructurar + T</i>	2
<i>Analizar</i> (2) <i>Estudiar</i> (2)			<i>Analizar/estudiar + T</i>	2
		<i>Tipos de</i> (2)	<i>Tipificar + T</i>	2

La question est maintenant de savoir si l'on peut considérer toutes les séquences sélectionnées comme des UF. Il nous semble que les collocations les plus « phraséologiques » sont celles dont le terme est le sujet soit d'un verbe événementiel intransitif à signification spécialisée (par exemple, *empresa + quebrar*), soit d'une expression spécialisée, sémantiquement opaque à première vue, comme, par exemple, *empresa + suspender sus/los pagos* ou *empresa + cotizar en bolsa*. Par contre, les séquences dans lesquelles le terme exerce la fonction de sujet ou de complément d'un verbe transitif à signification non ou peu spécialisée nous semblent s'approcher plus des unités discursives récurrentes, peut-être parce que les nombreuses possibilités de combinaisons avec d'autres termes et non-termes rendent la combinaison moins figée. Cette hypothèse doit être examinée plus amplement, mais il est clair qu'il existe un *continuum* de « moins phraséologique » vers « plus phraséologique ».

4. CONCLUSIONS

Dans cette tentative de formuler quelques prudentes conclusions, tant les résultats relatifs aux termes composés que ceux relatifs aux UF sont évalués en fonction de leur utilité didactique. Tout d'abord il convient insister sur l'importance de l'enseignement des TS, étant donné qu'ils constituent une des caractéristiques principales du langage spécialisé. Rappelons que 82,26 % du total des unités lexicales composées dans le *corpus* sont des termes (économiques ou auxiliaires) et que 66,19 % des termes économiques et 45,66 % des termes auxiliaires sont composés. Bien qu'il soit nécessaire de relativiser ces chiffres en taux de couverture, soit 29,05 % en termes économiques et 20,42 % en termes auxiliaires, les termes simples sont donc beaucoup plus fréquents, l'étude et l'analyse de la composition des TS doivent obligatoirement faire partie d'un cours de langue étrangère spécialisée afin que les apprenants puissent mieux les reconnaître, comprendre/en induire le sens, retenir et reproduire (Vangehuchten 2005).

Quant aux UF, nous pouvons conclure que le processus d'identification reste précaire. Voilà pourquoi il est impossible de proposer, sur la base de cette étude, des conseils didactiques, et la conclusion principale doit être qu'il est indispensable de continuer les recherches dans ce domaine. C'est précisément l'absence d'études phraséologiques plus poussées, tant descriptives (à base de *corpus*) qu'expérimentales (axées sur le processus psycholinguistique d'acquisition), qui fait que les auteurs Jones et Haywood (2004 : 276) concluent que « [...] *there is as yet no proven methodology for the teaching of formulaic sequences* ».

BIBLIOGRAPHIE

- Alvar Ezquerro (M.), 1993 : *La formación de palabras en español*, Madrid, Arco/Libros.
- Antia (B. E.), 2000 : *Terminology and language planning*, Amsterdam/Philadelphia, John Benjamins.
- Béjoint (H.) et Thoiron (Ph.), 1992 : « Macrostructure et microstructure dans un dictionnaire de collocations en langue de spécialité » dans *Terminologie et Traduction*, 2/3, pp. 513-522.

- Bevilacqua (C.), 2001 : « Unidades fraseológicas especializadas : elementos para su identificación y descripción » dans *La terminología científico-técnica : reconocimiento, análisis y extracción de información formal y semántica*, Barcelona, IULA (Universitat Pompeu Fabra).
- Biber (D.), 1993 : « Co-occurrence patterns among collocations : a tool for corpus-based lexical knowledge acquisition » dans *Computational Linguistics*, 19, 3, pp. 531-538.
- Cabré (M.T.), Estopà (R.) et Lorente (M.), 1996 : « Terminología y fraseología » dans *Actas del V Simposio Iberoamericano de Terminología Riterm*, publicación en Internet : <http://www.unilat.org/dtil/MEXICO/cabreloe.html>.
- Cabré (M. T.), 1999 : *La Terminología. Representación y comunicación*, Barcelona, IULA (Universitat Pompeu Fabra).
- Cabré (M. T.), 2001 : « Sumario de principios que configuran la nueva propuesta teórica » dans Cabré (M.T.) et Feliu (J.), édit., *La Terminología científico-técnica : reconocimiento, análisis y extracción de información formal y semántica*, Barcelona, IULA (Universitat Pompeu Fabra).
- Church (K. W.) et Hanks (P.), 1990 : « Word association norms, mutual information, and lexicography » dans *Computational Linguistics*, 16, 1, pp. 22-29.
- Church (K. W.) et al., 1994 : « Lexical substitutability » dans Atkins (B.T.S.) et Zampolli (A.), édit., *Computational approaches to the lexicon*, Oxford, Oxford University Press, pp. 153-177.
- Ellis (N. C.), 2002 : « Frequency effects on language acquisition » dans *Studies in second language acquisition*, 24, 2, pp. 143-188.
- Jones (M.) et Haywood (S.), 2004 : « Facilitating the acquisition of formulaic sequences : an exploratory study in an EAP context » dans Schmitt (N.), édit., *Formulaic sequences*, Amsterdam/Philadelphia, John Benjamins.
- Kennedy (G.), 1998 : *An Introduction to Corpus Linguistics*, Harlow, Addison Wesley Longman.
- Lewis (M.), 1993 : *The Lexical Approach*, Hove, Language Teaching Publications.
- Myking (J.), 1989 : « Complex noun phrases as a problem of terminological practice » dans Laurén (C.) et Nordman (M.), édit., *Special Language : from humans thinking to thinking machines*, Clevedon/Philadelphia, Multilingual Matters, pp. 265-274.
- Nation (I. S. P.), 2001 : *Learning vocabulary in another language*, Cambridge, Cambridge University Press.
- Pérez Gorostegui (E.), 1997 : *Introducción a la administración de empresas*, Madrid, Editorial Centro de Estudios Ramón Areces.
- Read (J.) et Nation (P.), 2004 : « Measurement of formulaic sequences » dans Schmitt (N.), édit., *Formulaic sequences*, Amsterdam/Philadelphia, John Benjamins.
- Schmitt (N.) et Celce-Murcia (M.), 2002 : « An overview of Applied Linguistics » dans Schmitt (N.), édit., *An introduction to Applied Linguistics*, Arnold, London, pp. 1-16.
- Schmitt (N.), édit., 2004 : *Formulaic sequences*, Amsterdam/Philadelphia, John Benjamins.
- Schmitt (N.) et al., 2004 : « Knowledge and acquisition of formulaic sequences » dans Schmitt (N.), édit., *Formulaic sequences*, Amsterdam/Philadelphia, John Benjamins.
- Smadja (F.), 1991 : « Macrocoding the lexicon with co-occurrence knowledge » dans Zernik (U.), édit., *Lexical acquisition : exploiting on-line resources to build a lexicon*, Hillsdale/New Jersey/Hove/London, Lawrence Erlbaum Associates, pp. 165-189.
- Smadja (F.), 1993 : « Retrieving collocations from text : Xtract » dans *Computational linguistics*, 19, 1, pp. 143-177.
- Stubbs (M.), 2001 : *Words and phrases. Corpus studies of lexical semantics*, Oxford/Massachusetts, Blackwell Publishers.
- Tercedor Sánchez (M.), 1999 : *La fraseología en el lenguaje biomédico : análisis desde las necesidades del traductor*, Estudios de Lingüística Española, 6, publicación en Internet : <http://elies.rediris.es/>.
- Thomas (P.), 1993 : « Choosing headwords from language-for-special-purposes (LSP) collocations for entry into a terminology data bank (term bank) » dans Sonneveld (H. B.) et Loenig (K.), édit., *Terminology. Applications in interdisciplinary communication*, Amsterdam/Philadelphia, John Benjamins, pp. 43-68.

- Vangehuchten (L.), 2005 : *El léxico del discurso económico empresarial : identificación, selección y enseñanza en Español como Lengua Extranjera con Fines Específicos*, Iberoamericana, Madrid.
- Wray (A.), 2002 : *Formulaic language and the lexicon*, Cambridge, Cambridge University Press.
- Wright (S. E.), 1997 : « Term selection : the initial phase of terminology management » dans Wright (S.E.) et Budin (G.), édit., *Handbook of Terminology Management*, vol. I, Amsterdam/Philadelphia, John Benjamins, pp. 13-23.

LES RÔLES DE LA SITUATION ET DU CONTEXTE DANS LES TECHNOLECTES BILINGUES FRANÇAIS – ARABE

Leila Messaoudi

UFR langage et société de sociolinguistique appliquée
département de langue et littérature françaises
Université Ibn Tofail
Kénitra, Maroc

Résumé : Deux volets sont abordés dans ce texte :

le premier est relatif au technolecte et à l'examen du lien à la situation via la relation au référent, en proposant une typologie (liens de type iconique, indiciel ou symbolique) ;

le second volet traite du technolecte et l'importance du contexte i.e. de l'environnement linguistique. L'exemple choisi est celui du technolecte linguistique.

Des questions épineuses se posent lors du passage de langue à langue et des télescopes conceptuels s'opèrent.

Mots-clés : *technolecte bilingue, contexte, situation, lien iconique, lien indiciel, lien symbolique*

Le rôle de la situation et du contexte dans les technolectes bilingues est une vaste problématique dont nous n'aborderons ici que quelques aspects.

Ne prétendant pas à l'exhaustivité, cette réflexion se fera selon deux volets :

– le premier consistera à essayer d'identifier le rôle de la situation via le lien au référent, à travers quelques exemples puisés essentiellement dans des domaines techniques qui viendront illustrer les types de relations repérées ;

– le second portera sur le rôle du contexte linguistique et le lien au paradigme conceptuel en prenant un exemple dans la terminologie grammaticale.

Auparavant, nous tenterons d'explicitier ce que nous entendons par les termes de contexte et de situation.

1. PRÉLIMINAIRES

1.1. Contexte et situation

Les deux termes *contexte* et *situation* présentent une utilisation peu concordante dans les travaux linguistiques et sont parfois employés comme synonymes.

Si le terme de *situation* ne prête pas à confusion puisqu'il renvoie généralement à l'environnement extralinguistique, il n'en est pas de même pour le terme de *contexte*, qui peut référer aussi bien à l'environnement linguistique syntagmatique *i.e.* les unités linguistiques précédant et suivant l'unité en question (dans le cadre de l'énoncé, la phrase ou le texte) – qu'à l'environnement non linguistique c'est-à-dire à la situation. Le recours à la syntagmatisation adjectivale permet parfois de lever l'ambiguïté en usant par exemple de « contexte linguistique » et « contexte extralinguistique » (Labov 1972) ou de « contexte verbal » et de « contexte situationnel » (Dubois *et al.* 1994).

Évitant de donner une définition qui ne pourra être que réductrice, nous convenons d'utiliser le mot « contexte » pour désigner l'environnement linguistique et celui de « situation » pour l'environnement extralinguistique. Puis, nous emploierons le terme de « microcontexte » pour renvoyer à l'environnement immédiat *i.e.* les mots précédant et suivant l'unité considérée ; tandis que le terme de « macrocontexte » servira à désigner un environnement large comprenant la phrase, le paragraphe, le discours.

1.2. Contexte, situation et technolecte

Depuis que ses fondateurs (notamment Wuster et à sa suite Felber) l'ont érigée en science, au carrefour de l'épistémologie, la logique et la sémantique, la terminologie a fait l'objet de questionnements multiples relatifs à la méthodologie et surtout à la construction de systèmes conceptuels à partir des notions et des termes (unités terminologiques simples ou complexes) qui en sont les supports. La terminologie a posé que des relations de type paradigmatique relient les notions entre elles, dans un ordre hiérarchique précis (hyperonymie, hyponymie, etc.), en dehors de tout contexte ou situation. Autrement dit, il faudrait mettre les termes hors circuit *i.e.* hors interaction communicationnelle, hors travail, pour mieux les décortiquer *in vitro*.

Le technolecte (défini, entre autres, dans Messaoudi 1990, 2000, 2001, 2003), fonctionne comme la terminologie, avec la différence que les unités de cette dernière ont le statut de « terme » et sont normalisés tandis que les unités du technolecte peuvent être normalisés ou non et surtout, elles intègrent les phrasèmes, constructions phraséologiques souvent évitées par les terminologues. De même que la terminologie, le technolecte d'un domaine ou d'une sphère d'activités a pour objet des systèmes conceptuels. Comme en terminologie, le sens spécialisé semble affecter les unités ayant la capacité d'être décontextualisées. Ce sens est valide, en principe, indépendamment des contextes et des situations, à des fins de scientificité et d'universalité. Cette optique quelque peu restrictive l'est délibérément, et c'est à ce prix, en procédant en quelque sorte à une momification des unités, que l'on pourra les analyser. Pourtant, la signification d'une unité ne peut être saisie qu'en fonction de son lien à la situation — tout particulièrement au référent qui est omniprésent dans la dénomination technolecte.

Notons en passant que depuis les années 1980, l'acception du terme technolecte a connu une certaine extension et a tendance à remplacer l'usage fort contesté de « langue de spécialité ».

Il va de soi que d'autres éléments de la situation pourraient être fort éclairants pour l'étude du technolecte, notamment son fonctionnement sociolinguistique et ses usages.

Mais dans le cadre restreint de cet article, sachant que la situation de communication au sein de laquelle se déroule l'acte technolectal peut être très diverse (scolaire, didactique, scientifique, professionnel, etc.), il ne sera pas tenu compte des aspects situationnels et le seul élément pris en compte sera le référent en raison de son importance pour le technolecte. La question qui se pose est : Quel lien au référent dans la dénomination technolectale ?

2. TECHNOLECTE ET SITUATION : LE LIEN AU RÉFÉRENT

Dans la théorie du signe linguistique, Saussure pose que le lien entre le signifiant et le signifié est immotivé. Mais en observant le fonctionnement de certaines unités technolectales, on peut constater que ce lien n'est pas toujours arbitraire et qu'une motivation s'instaure à travers le renvoi au référent. Or, le sens de la motivation chez Saussure est essentiellement intralinguistique (Saussure 1974 : 182-184 ; Gadet 1987 : 101-104). En réalité, la visée sémiotique de Saussure se révèle peu pertinente pour l'approche des technolectes et de la terminologie de façon générale (Lerat 1994 : 37).

Pour le technolecte, l'évacuation de l'extralinguistique est quasi impossible tant le lien avec le référent (chose ou objet) est non seulement nécessaire (pour reprendre la formulation de Benveniste) mais obligatoire.

Nous inspirant de la trichotomie de C. S. Peirce, tout en opérant les transferts conceptuels utiles pour notre argumentation, nous pourrions avancer l'idée que les trois types de signes dégagés par Peirce : indice, icône et symbole sont supportés chacun par un mode de désignation impliquant un lien particulier au référent, lorsqu'ils sont appliqués à des unités linguistiques de type technolectal.

Notre hypothèse est que le mode de désignation dans le technolecte est motivé par la nature du lien instauré avec le référent. Comment qualifier ce lien ?

2.1. Le lien de type iconique

C. S. Peirce définit l'icône comme suit : « n'importe quoi, qualité, individu existant ou loi, est l'icône de quelque chose pourvu qu'il ressemble à cette chose et soit utilisé comme signe de cette chose » (Deledalle 1979 : 42). L'icône est à interpréter comme l'image reproduisant fidèlement un objet, créant ainsi une relation d'identité entre eux. Le lien iconique référerait au mode où la motivation est à son degré le plus fort et où le processus de l'imitation du référent tend vers la représentation idéale qui est celle de l'image. Les onomatopées sont des exemples de ce mode de désignation puisque leurs signifiants tentent de copier ou reproduire les sons proférés par les animaux ou bien ceux de la nature, etc.

Ce souci de reproduction de la forme du référent ou de l'un de ses traits peut caractériser certaines unités technolectales. Toutefois, à l'exemple des onomatopées, ces unités ne sont pas nombreuses.

Dans le technolecte, ce procédé est utilisé dans les exemples suivants :

- le T (règle en T)
- antenne en T
- virage en Z
- crochet en S

Il est intéressant de noter que dans ce procédé, le mode de désignation a consisté à choisir parmi les graphèmes de l'alphabet latin ceux qui reproduisent la forme du référent en question ou ceux qui s'en rapprochent le plus.

Que se passe-t-il dans les technolectes bilingues français-arabe ? Comment ce mode est-il rendu ?

Voici quelques exemples :

- antenne en T	<i>Hawà'iyah calà fakl T</i>
- crochet en S	<i>Kullàb bi fakl S</i>
- le T (té)	<i>Mistara bi fakl T</i>
- virage en Z	<i>Muncataf bi fakl Z</i>

Si le souci de reproduire ou d'imiter la forme du référent est présent dans la langue source (le français), il l'est aussi dans la langue cible (l'arabe) : le traducteur n'a pas hésité à retenir le caractère latin puisque ce n'est pas le graphème membre de l'alphabet qui est visé mais sa forme sémiotique qui importe. Le principe de l'imitation étant central, les caractères T, S, Z, etc. repris dans la traduction arabe, faute d'équivalents, fonctionnent bien comme un mode de désignation de type iconique.

2.2. Le lien de type indiciel

Deledalle signale que pour Peirce « l'indice renvoie à l'objet qu'il dénote parce qu'il est réellement affecté par cet objet » (*op. cit.* p. 74). Il note la différence entre l'indice et l'icone en disant : « alors que l'icone est liée à son objet par la ressemblance, l'indice l'est par la contiguïté » (*ibid.*, p. 75).

Nous qualifierons d'indiciel, le mode de désignation qui consiste à renvoyer au référent par l'une ou plusieurs de ses caractéristiques. À la différence du mode iconique qui tente de reproduire l'ensemble du référent (chose ou objet), sa forme globale par exemple dans « antenne en T », le mode indiciel fait appel à une partie du référent, pour fournir une précision supplémentaire. Ainsi, se tisse un lien de « contiguïté » exprimé sur le plan syntagmatique par la coexistence d'éléments selon des types syntaxiques précis. Par exemple : *herse à dents*, *herse à dents souples*, *herse à dents rigides*. Dans ces trois unités technolectes, il est fait référence, chaque fois, à un trait de l'objet. Qu'est-ce qui différencie les herses ? L'une est à dents souples, l'autre à dents rigides. C'est à une partie

et non au tout que ce procédé fait appel. Il est ainsi de type métonymique à la différence du mode iconique qui serait plutôt de type équatif si l'on considère que dans « antenne en T », antenne est égale à T.

Comment se trouvent rendues ces unités en arabe ? Les exemples cités ci-dessus, à titre d'illustration, seront traduits en ayant recours au calque par la traduction mot-à-mot ou littérale :

-	Herse/à/dents/flexibles
-	<i>mu ftun/bi/asnànin/marinah</i>
-	herse à dents rigides
-	<i>mu ftun/bi/asnànin/jàsi'ah</i>

2.3. Le lien de type symbolique

Selon Peirce « Le symbole est un signe qui renvoie à l'objet qu'il dénote en vertu d'une loi » (*op. cit.* p. 76) note G. Deledalle qui ajoute : « Tout mot, tout signe conventionnel est un symbole » (*ibid.*, p. 76)

Les unités du technolecte dans ce troisième cas de figure fonctionnent comme l'ensemble des mots de la langue *i.e.* de manière tout à fait conventionnelle, sans rechercher un lien explicitement exprimé avec le référent.

En fait, les signes mots se substituent métaphoriquement aux référents, sans chercher à les reproduire tout ou partie, comme c'est le cas du mode indiciel ou du mode iconique. Il s'agit là d'une relation de substitution sans motivation *i.e.* purement arbitraire du point de vue du lien avec le référent, qui caractérise tout signe du langage spécialisé ou non.

Vraisemblablement, la part de la motivation augmente si l'on part du symbole vers l'indice puis vers l'icone. Proportionnellement, le nombre d'occurrences motivées par le lien au référent dans le technolecte est exprimé le plus souvent par le mode indiciel et baisse progressivement pour atteindre le niveau le plus bas dans le mode iconique. On peut postuler que cette progression doit être la même pour la langue générale, à la différence que dans le technolecte, le mode de désignation indiciel doit être plus fréquent.

Ainsi, pour certaines unités technolectes, une large part de motivation via le lien avec le référent est perceptible dans le cas des modes de désignation iconique et indiciel tandis que le reste des unités technolectes fonctionne vis-à-vis du lien au référent, pratiquement de la même manière que les unités de la langue générale.

Ces suppositions pourraient être vérifiées sur de larges *corpus* afin de voir si vraiment l'élément extralinguistique qu'est le référent est plus « présent » dans le technolecte que dans la langue générale ou s'il s'agit, en fait, de la mise en œuvre des mêmes modes iconique, indiciel et symbolique et que seules les proportions peuvent varier.

Ceci reste à démontrer, l'important pour nous est d'affirmer que le référent, en tant qu'élément externe, est pris en compte tout aussi bien dans la dénomination de type technolectal que dans celle de la langue en général et que les mêmes processus cognitifs sont mobilisés. En revanche, pour les technolectes bilingues considérés, ces modes de désignation s'avèrent compliqués, puisque la relation au référent se fait par l'intermédiaire de la langue L1 avant d'être transposée en L2.

Le comportement du traducteur consiste, dans la majeure partie des cas, à opérer des calques, de L1 vers L2, sans se soucier de savoir si tel appareil conceptuel ou telle ou telle représentation existent dans la langue cible. Pour prendre un exemple où le lien avec le référent est de type symbolique et même métaphorique, on peut citer dans le technolecte médical, « ichtyose » qui désigne « un état particulier de la peau, qui est sèche et recouverte de squames, semblables aux écailles de poisson » et qui a été rendu en arabe par « dà'u ssamak » littéralement « maladie de poisson » ! Traduction pas très heureuse dans la mesure où un locuteur arabophone peut penser que cette dénomination, calquée littéralement sur l'item de la langue source, désigne une maladie qui affecte le poisson et non la peau d'un être humain ! Ce risque n'interviendrait-il pas en français ? Le recours aux formants savants (du grec *ikhthus* + *isis* « mal ») contribue-t-il à diminuer l'opacité et la confusion ? Peut-être, mais dans tous les cas, on gagnerait en transparence notionnelle si l'on prenait en compte le contexte à la fois au niveau micro et macro.

3. TECHNOLECTES ET CONTEXTE LINGUISTIQUE

L'on constate à travers les pratiques terminologiques et les recherches sur les technolectes, que les termes requièrent bien souvent un environnement linguistique, représenté par un microcontexte et parfois un macrocontexte pour que leur contenu soit moins opaque. Les exemples sont nombreux où le terme en tant que tel a besoin de l'éclairage du contexte pour atteindre un maximum de transparence.

Pour ce qui nous concerne, nous allons nous interroger sur la pertinence du contexte pour le traitement du terme dans le métalangage grammatical. Nous avons construit un *corpus* autour d'un exemple, celui de *harf*.

3.1. L'exemple de *harf* dans le technolecte linguistique

Pour la constitution du *corpus* de base, les données sont puisées du *Dictionnaire unifié de linguistique* (désormais *DUL*), dans ses deux versions (1989 et 2002). Les entrées, suivies de l'équivalent en français, sont regroupées selon leur appartenance à tel ou tel champ linguistique (phonologie, morphosyntaxique) (voir tableau ci-dessous).

	DUL 1989		DUL 2002	
Domaine	Entrées	Équivalent Français	Entrées	Équivalent Français
Phonologie	harfun ‘adlaq (=ma’ic)	liquide sonant	musawwit ma’ic	sonant liquide
	harfu ttafa ʃi (= Sawtun mutaffa ʃʃin)	chuintante	∅	∅
	harfun habsyun basit (=harfun ʃadid basit)	occlusive simple	insidàdi ʃadid	occlusif
	harf rixw	consonne spirante	ihtikàki	spirant
	harf ʃadid	stop	∅	∅
	harf Sawtiy	phonème	sawtiyyah	phonème
	harfu llin	– glide – semi-consonne – semi voyelle	cillah	glide
Morpho-syntaxe	harfun ‘asliy	radical	‘asl	radical
	harfu lcaTf	coordonnant	harfu lcaTf	coordonnant
	harfu jarr	préposition	harf	préposition
	harfu jarrin mu’axxar	préposition en fin de phrase	∅	∅

Remarques :

On peut noter que dans le *DUL* (1989), « harf » apparaît dans des unités appartenant aux domaines de la phonologie et de la morphosyntaxe alors que pour le *DUL* 2002, il n’apparaît que pour le seul domaine de la morphosyntaxe.

Dans la version de 2002, harf est retenu avec le sens de préposition. Il est cité aussi dans le sens usuel de « lettre de l’alphabet, caractère », *harf* = graphème. Dans les composés, il est employé pour rendre un phénomène inexistant dans la graphie arabe, celui des majuscules et des minuscules, rendu respectivement par *harf kabir* et *harf ‘adi*.

Dans la version de 1989, *harf* n'apparaît jamais seul et est toujours syntagmatisé.

Pourquoi cette différence de traitement de *harf* dans les deux versions ?

3.2. « harf » dans le métalangage grammatical arabe

En comparant les items repérés (*harf* seul ou en composé *harf* + substantif, *harf* + adjectif), il faudrait essayer de cerner leur place dans le système conceptuel grammatical.

Pour ce faire, on peut se référer au traitement de l'item *harf* dans *Lisàn al Arab* qui présente le terme au sein de *corpus* (citations essentiellement).

On peut aussi procéder à un dépouillement exhaustif de *harf* en microcontexte, dans la tradition grammaticale ancienne. Par exemple, dans le *kitàb* de Sibawaihi, *harf* apparaît au singulier dans des syntagmes comme *harfu nafy* (particule de négation), *harfu l'istifhàm* (particule d'interrogation), *harfu jarr* (préposition), *harfu ttacajjub* (particule d'exclamation) mais aussi au pluriel dans *huruf al carabiyya*, traduits dans ce contexte par « sons de la langue arabe », dont il va présenter les *maxàrij i.e.* les points d'articulation sur le plan phonétique.

Une approche qui regrouperait toutes les occurrences de *harf* en microcontexte mais aussi en macrocontexte (par exemple, l'étude des points d'articulation des *huruf* figure chez Sibawaihi dans la partie consacrée au '*idRam* (l'assimilation)). Ce qui relève d'une conception de la langue et de ses composants et d'une catégorisation propre. Cette catégorisation sera-t-elle reprise par le traducteur ?

3.3. Le passage de langue à langue et le contexte

Le passage de langue à langue, en l'occurrence du français vers l'arabe, dans le cas des lexiques spécialisés à métalangue spécifique, appartenant à des systèmes conceptuels différents, permet-il de faire l'économie du contexte ?

Par exemple, certaines notions de la grammaire arabe chez Sibawaihi se trouvent transposées dans les traductions par des termes de la linguistique saussurienne et post-saussurienne, certaines équivalences sont posées sans tenir compte de l'appareil conceptuel dans l'une et l'autre pensée. Des projections d'une conception mènent parfois à des erreurs d'interprétation et de traduction.

Revenons au traitement de *harf* dans *DUL* 1989 : on s'aperçoit que la catégorisation suivie serait, toutes proportions gardées, celle de Sibawaihi pour qui le mot *harf* apparaît aussi bien au niveau morphologique (*sarf*) qu'à celui de l'étude du son (*sawt*). Cette catégorisation originelle a été transposée, non sans subir quelques réajustements pour la faire correspondre aux domaines de la linguistique où *harf* se retrouve équivalent à son, phonème, préposition, adverbe, conjonction, etc. Il semble que *harf* y soit pris au sens de « petite unité » pouvant appartenir au niveau phonologique, morphosyntaxique ou sémantique. Ceci n'est pas sans rappeler « la plus petite unité distinctive ou significative » de Martinet. Mais cette interprétation existe-t-elle dans la pensée de Sibawaihi pour qui le *harf* est défini comme : « *harfun jà'a limacna, laysa bi'ismin wa là fiçl* » (*harf* correspond à un sens et (il) n'est ni un nom ni un verbe) ?

Une étude plus approfondie prenant en considération les différents contextes dans lesquels apparaît « harf » chez Sibawaihi apporterait quelques éléments de réponse à cette question. En attendant, l'emploi de *harf* a conduit les auteurs du *DUL* 1989 à des constructions produisant des dénominations curieuses telles que : *Cilmu lhurufi lmuqattacah* pour traduire « phonologie segmentale » ou bien *huruf muqattacah* pour « phonèmes segmentaux » qui semblent bien éloignés de la conception que se faisait Sibawaihi du *harf* !

Dans le *DUL* 2002, le traitement de *harf* est différent. Les auteurs n'ont pas repris la catégorisation dégagée du *Kitàb* de Sibawaihi mais ont essayé de transférer la conceptualisation moderne de la linguistique, telle qu'exprimée via la langue source, vers la langue cible et n'ont pas essayé de forcer le contenu de *harf* en opérant des projections de sens modernes. Ils se sont contentés de retenir les dénominations admises en morphologie arabe et qui continuent encore d'être usitées actuellement dans le métalangage grammatical pour désigner des éléments comme *harfu lcatf* « le coordonnant », *harf* « la préposition ». Pour ce qui concerne les aspects phonétiques, les auteurs ont eu recours à la terminologie arabe moderne, en usage dans les publications linguistiques comme celles de Fassi Fihri, Mseddi, Seghrouchni entre autres, et ont évité l'emploi de *harf*. Mais ce n'est que grâce à une exploration textuelle à travers le *Kitàb* de Sibawaihi que cette prise de conscience du système conceptuel auquel appartient *harf* a pu s'imposer aux auteurs et les a conduits à prendre une décision lexicographique différente de celle de leurs prédécesseurs.

4. CONCLUSION

Au terme de cette réflexion, on constate que :

- l'examen des unités technolectoales en relation avec la situation permet de déceler des liens avec le référent, exprimés par des modes de désignation particuliers (iconique, indiciel et symbolique) qui ne sauraient être ignorés dans l'approche des technolectes bilingues ;
- la vigilance du traducteur ne saurait éluder les contextes dans lesquels se construit le sens des unités. L'exemple de *harf* a été pris à titre indicatif. D'autres exemples de la mise en communication de deux systèmes conceptuels sont nombreux, on peut citer celui de l'expression du temps telle qu'elle est exprimée par les grammairiens arabes et celle des arabisants européens (Messaoudi 1985).
- Les relations conceptuelles passent par des catégorisations qui semblent déterminées par des découpages prélinguistiques qu'il s'agit de repérer et de prendre en compte. Ces découpages paraissent aisément décelables dans les technolectes techniques à mode de désignation iconique ou indiciel, où le référent objet revêt une certaine stabilité. En revanche, ils le sont plus difficilement dans le mode symbolique qui semble caractériser le technolecte grammatical et linguistique, en raison de l'absence d'un référent stable. D'où la nécessité de recourir aux contextes aussi bien en langue L1 que L2 pour opérer un rapprochement si nécessaire ou du moins pour parvenir à des études comparatives pouvant mener à des recherches approfondies sur des unités qui baignent dans un environnement dont elles tirent toute leur substance.

À cet égard, une étude exploratoire à l'aide de l'outil informatique serait la bienvenue et permettrait de repérer les contextes pouvant aider à cerner les différentes significations de tel ou tel item, selon qu'il apparaît au niveau micro *i.e.* environnement immédiat ou macro *i.e.* environnement large, dans tel ou tel chapitre ou dans tel ou tel passage textuel (phrase, paragraphe).

Des *corpus* bilingues alignés, mettant en relation les concepts de la langue source et ceux de la langue cible tels qu'ils apparaissent dans leur environnement originel (mais non traduits) seront réellement utiles.

On ne peut donc que souscrire à l'idée de construire des bases de données textuelles et « l'avènement de la terminologie textuelle » (voir entre autres Condamines 2005) est révélateur de l'intérêt de ces outils.

De plus, les progrès en TAL et dans le développement de logiciels susceptibles de repérer des réseaux grâce aux « marqueurs de relations conceptuelles », ont nettement fait avancer la recherche en linguistique du *corpus* (voir entre autres Condamines 2000).

Par ailleurs, le recours aux éléments extralinguistiques dans un cadre sociolinguistique, voire socioterminologique (Gaudin 2005) ne participerait-il pas à une meilleure appréhension des technolectes bilingues ?

BIBLIOGRAPHIE

- Bureau permanent d'arabisation, 1989 : *Dictionnaire unifié de linguistique*, Tunis, Publication de l'ALECSO.
- Bureau permanent d'arabisation, 2002 : *Dictionnaire unifié de linguistique*, Tunis, Publication de l'ALECSO.
- Condamines (A.), 2005 : « Linguistique de *corpus* et terminologie », dans *Langages*, 157, Paris, pp. 36-47.
- Deledalle (G.), 1979 : *Théorie et Pratique du signe*. Paris, Payot.
- Dubois *et al.*, 1994 : *Dictionnaire de linguistique*, Paris, PUF.
- Fassi Fihri, 1987 : *Al lisàniyyàt al arabiyya*, Casablanca, Éd. Toubkal.
- Gadet (F.), 1987 : *Saussure. Une Science de la langue*, Paris, PUF.
- Gaudin (F.), 2005 : « La socioterminologie », dans *Langages*, 157, Paris, pp. 80-92.
- Labov (W.), 1972 : *Sociolinguistique*, Paris, Minuit.
- Lerat (P.), 1994 : *Les Langues spécialisées*, Paris, PUF.
- Mejri (S.), 2003 : « La traduction linguistique : problème terminologique ou construction conceptuelle ? » dans *Traduire la langue, Traduire la culture*, Paris, Maisonneuve et Larose.
- Messaoudi (L.), 1985 : *Temps et Aspect. Approche de la phrase simple en arabe écrit*, Paris, Geuthner.
- Messaoudi (L.), 1986 : « L'utilisation de la base de données *Lexar* de l'IERA dans la recherche des emprunts au français dans la terminologie scientifique arabe », dans *Dialogue entre la langue arabe et la langue française*, Paris, CILF.
- Messaoudi (L.), 1990 : *Des technolectes. Approche linguistique. Application à l'arabe standard*, doctorat d'état, Paris V, Lille, Centre de reproduction des thèses.
- Messaoudi (L.), 1995 : *Traduction et Linguistique. Le Cas des technolectes*, Publications de la faculté des lettres et des sciences humaines, Rabat.
- Messaoudi (L.), 1998 : « Technolectes bilingues (français-arabe) et modes de dénomination » in *Revue du Ceres*, Tunis.

- Messaoudi (L.), 2000 : « Opacité et transparence dans les technolectes bilingues (français – arabe) », dans *Meta*, XVI, 3, Presses de l'Université de Montréal, pp. 424-436.
- Messaoudi (L.), 2002 : « Le technolecte et les ressources linguistiques », revue *Langage et Société*, Paris.
- Messaoudi (L.), 2004 : « Les technolectes au Maroc. Fonctionnement et tendances d'évolution », dans *Trames de langues*, Paris, Maisonneuve et Larose, pp. 439-455.
- Mseddi (A.), 1984 : *Dictionnaire de linguistique français-arabe, arabe-français*, Tunis, Maison arabe du livre.
- Rastier (F.), 1995 : *Le Terme : entre ontologie et linguistique*, la Banque des mots, Paris, pp. 36-65.
- Saussure (F.), 1974 : *Cours de linguistique générale*, Paris, Payot.
- Seghrouchni (D.), 1987 : *Assiwàta ttawliḍiyah*, Casablanca, Toubkal.
- Sibawaihi, 1967 : *Al kitāb*, Beyrouth, publications Mu'asasat l'alami.

LA TERMINOLOGIE DES SCIENCES DE GESTION EN CONTEXTE : TERME OU MOT ?

Soumaya Mejri

sciences de gestion

Institut d'administration des entreprises

Université de Caen, Basse-Normandie

Caen, France

soumayamejri@yahoo.fr

Résumé : *Par une analyse de discours nous montrons que le contexte académique se caractérise par une consistance terminologique et le contexte général se caractérise par une dissolution parfaite des concepts dans le discours. Dans le discours des praticiens, le statut de la même unité lexicale oscille entre mot et terme. Cette analyse montre que le statut de la même unité lexicale varie en fonction de la nature du contexte discursif. Autrement dit c'est le contexte qui détermine le statut des unités lexicales.*

Mots-clés : *sciences de gestion, terme, contexte, prospective, prévision et planification*

Dans ce travail, nous nous proposons d'analyser l'impact du contexte sur les différentes unités lexicales. Pour effectuer un tel travail, nous avons eu recours au vocabulaire des gestionnaires qui se caractérise par la diversification des concepts et la richesse des notions. On change de discours autant de fois que d'interlocuteurs. Cette pluralité de discours présente une variété de contextes distincts et fait de la gestion un bon exemple pour étudier la relation entre mot, terme et contexte.

En se basant sur la terminologie des gestionnaires et leurs discours, on va essayer d'apporter quelques éléments de réponse à la problématique suivante :

En quoi consiste la relation entre mot, terme et contexte ? Et comment le contexte explique l'usage des mots ou des termes ?

Pour traiter cette problématique, nous avons choisi trois concepts fondamentaux en matière de « prospective d'entreprise » et essayé de les définir à partir des dictionnaires spécialisés en sciences de gestion qui présentent des références incontestables dans la matière.

À partir de ces définitions, nous avons dégagé les éléments définitoires de chaque concept afin de mieux le cerner et d'en délimiter les contours. Finalement, nous avons étudié ces concepts dans les différents discours des gestionnaires afin de dégager la relation entre mot, terme et contexte.

En résumé, nous partons des définitions courantes des trois termes linguistiques *mot*, *terme* et *contexte* pour les appliquer aux *items* lexicaux choisis *prospective*, *prévision* et

planification dans des discours de gestion afin de vérifier l'hypothèse que le statut de la même unité lexicale varie en fonction de la nature du contexte discursif.

Pour reformuler l'hypothèse de notre travail, nous retenons :

Plus le discours est spécialisé, plus la consistance conceptuelle est grande et inversement.

Dans une première partie, nous fournirons dans ce qui suit les définitions des termes linguistiques, la structure conceptuelle des trois items lexicaux et la présentation de notre *corpus*. Dans une seconde partie, nous procéderons à l'analyse des fragments discursifs choisis.

1. DÉFINITIONS ET STRUCTURES CONCEPTUELLES

1.1. Définitions des termes linguistiques : terme, mot et contexte

Pour définir ces différents concepts linguistiques, nous nous sommes basée sur quelques dictionnaires de linguistique de référence et sur l'ouvrage de base de P. Lerat (1995).

1.1.1. Le terme :

C'est la « désignation d'une notion sous forme de lettres, de chiffres, de pictogrammes ou d'une combinaison quelconque de ces éléments » (p. 17). P. Lerat précise que les termes sont principalement des mots et groupes de mots (nominaux, adjectivaux, verbaux) sujets à des définitions conventionnelles. Le terme doit avoir un sens spécialisé et une syntagmatique restreinte. Cet auteur précise que pour reconnaître un terme de quelque nature que ce soit (mot radical, abréviation, dérivé, composé ou locution), nous pouvons nous baser sur quelques critères de reconnaissance :

- (1) Le terme appartient à une série morphologique de mots de sens spécialisé au sein de la langue considérée ;
- (2) Le terme a une syntagmatique restreinte (cooccurrences et commutations dans les limites d'un domaine spécialisé) ;
- (3) Le terme a des relations de dépendance par rapport à des unités présentant les caractéristiques 1 et 2.

Dans *Le Dictionnaire de linguistique* (1973), on précise qu'en syntaxe, le terme est un mot qui assume, dans une phrase, une fonction déterminée. Cette définition appuie les propos de P. Lerat disant que le terme est restreint par rapport au mot, il est monosémique et doit avoir un sens bien délimité.

1.1.2. Le mot :

Le mot est selon Meillet « L'association d'un sens donné à un ensemble de sons donné susceptible d'un emploi grammatical donné » (voir Lerat 1983 : 7, 1995).

En linguistique structurale, la notion de mot est souvent évitée en raison de son manque de rigueur. Elle intervient encore dans une opposition terme vs mot. « Terme » désigne ici

l'emploi monosémique (possédant une signification unique) qui sera fait d'une unité lexicale dans telle ou telle science, soucieuse d'établir une correspondance univoque entre ses concepts et les termes de sa nomenclature (par exemple, *rayon* est un terme scientifique de l'électrostatique, dans *rayon X*, *rayon gamma*, etc.). « Mot » désignera, dans cette opposition, l'unité lexicale du vocabulaire général, essentiellement polysémique (susceptible de significations variées). P. ex. : *rayon* dans *chef de rayon*, *rayon de soleil*, *roue à rayons*, etc.

1.1.3. Le contexte :

Dans *Le Dictionnaire de linguistique* (1973), le contexte est tout d'abord défini d'une manière générale comme l'environnement, c'est-à-dire les unités qui précèdent et qui suivent une unité déterminée. Ensuite, il a été défini comme l'ensemble des conditions sociales qui peuvent être prises en considération pour étudier les relations existant entre le comportement social et le comportement linguistique. Cet ensemble désigne souvent le contexte social d'utilisation de la langue. On dit aussi contexte situationnel. Donc le contexte présente l'entourage linguistique et non linguistique dans lequel une unité est produite ou reçue.

Dans *Le Dictionnaire de didactique des langues* (1976), au niveau de la « parole », le contexte désigne l'entourage linguistique d'une unité, c'est-à-dire l'ensemble des éléments réellement présents dans le texte au voisinage immédiat ou éloigné de l'unité considérée. Ces éléments qui conditionnent la présence, la forme, la fonction ou le sens de cette unité appartiennent au contexte pertinent. P. ex. : dans « Avez-vous une cuisinière ? Oui, j'en ai acheté une », *acheter* est pertinent parce qu'il sélectionne le sens de cuisinière.

Selon Jakobson (voir *Le Dictionnaire didactique des langues* 1976 : 123), le contexte est équivalent de référent (terme dont Jakobson dénonce l'ambiguïté). Dans sa terminologie, la fonction référentielle renvoie au contexte qui peut être extralinguistique ou linguistique.

Partant de ces différentes définitions, nous retenons les caractéristiques suivantes : le terme, puisqu'il a une contrepartie conceptuelle bien délimitée, ne peut avoir d'emploi approximatif que s'il est controversé (mis en discussion) ou s'il est employé dans un contexte non spécialisé.

- Au contraire, le mot, puisqu'il a en contrepartie un signifié, par définition flou, s'intègre dans des combinaisons lexicales de synonymie partielle (relations d'hyponymie, d'hyperonymie) ou de paraphrase.
- Le contexte se définit selon plusieurs paliers allant de la phrase au texte large. Les paramètres du contexte phrastique relèveraient des contraintes sémantico-syntaxiques, le contexte large de paramètres croisés (sémantique, lexique et syntaxe) et de critères extralinguistiques.

1.2. Structure conceptuelle des trois termes : prospective, prévision et planification

Il s'agit des termes-clés qui définissent le domaine de gestion prospective qui implique le futur dans la gestion de toute organisation. La prospective est une démarche globale de réflexion sur le devenir d'un système ou d'une organisation, comportant différents

scénarios de possibilités, dans le but de déceler certaines tendances lourdes dans les évolutions futures. « C'est un regard sur l'avenir destiné à éclairer l'action présente » (Godet 1992 : 32). Ce nouveau concept permet de cerner une nouvelle vision du futur, de promouvoir une nouvelle attitude face à l'avenir et de développer de nouvelles méthodes pour appréhender des environnements turbulents. La prospective peut être définie également comme une discipline qui étudie les causes techniques, scientifiques et sociales qui accélèrent l'évolution du monde moderne. Elle se caractérise par la distinction entre l'essentiel et l'accessoire, entre le durable et l'éphémère et entre le global et le spécialisé. Cette discipline a recours à plusieurs méthodes : des outils de modélisation, des méthodes de consultation d'experts, des techniques d'animation de groupe, des méthodes de simulation diverses et des méthodes de scénario. Elle se distingue des autres approches du futur (la prévision, la planification et la futurologie) car elle se libère de l'attitude contemplative pour aider davantage à l'action (éclairer les choix et ajuster les stratégies). Elle s'impose à cause de la rapidité des changements surtout technologiques et permet une action efficace sur les événements afin de préparer et atteindre l'avenir souhaitable. La dynamique de la prospective est certes fortement tournée vers l'avenir mais elle cherche ses racines dans le passé et dans le présent puisque son rôle est moins de prévoir l'avenir que de comprendre les mécanismes de cette articulation (passé, présent et futur) afin d'agir dans le présent.

Quant à la prévision, elle est une détermination chiffrée ou non d'un phénomène, d'une grandeur ou d'un ensemble de grandeurs relatifs à une période future. C'est une opinion formée par le raisonnement sur les choses futures. Il s'agit d'une démarche d'anticipation. La différence entre les termes « prévision » et « anticipation », qui sont souvent utilisés dans un sens voisin pour désigner la « scrutation » du futur par un acteur (attitude prévisionnelle ou anticipatrice) :

- *L'anticipation* revient à considérer un événement futur comme s'il s'était produit.
- *La prévision* consiste à se représenter à l'avance une chose probable. Elle peut être à court, moyen ou long terme. Elle peut être déterministe ou probabiliste.

Prévoir, c'est à la fois supputer l'avenir et le préparer ; c'est déjà agir (Fayol 1916). Il s'agit d'anticiper un futur probable en l'extrapolant à partir d'une étude approfondie des événements passés ou présents. C'est une affaire de vigilance, de calculs et de dispositions à prendre. Essayer de prévoir l'avenir, c'est pour savoir quoi faire, pour se développer à moindre risque ou pour se protéger (contre des dysfonctionnements, des risques ou des crises). Dans une démarche de prévision, on peut se hasarder à faire des pronostics, c'est-à-dire des affirmations hypothétiques sur l'avenir en les appuyant ou non sur un processus de diagnostic. En économie, c'est l'ensemble d'études mettant en œuvre les données de la statistique, les théories économiques, les conditions non économiques, etc.

La planification, dans le *Lexique de gestion* (2003), est définie comme la mise en ordre dans le temps des objectifs de gestion des domaines d'activité. Elle suit la segmentation qui est la phase d'analyse et de décision initiale (que fait-on ?). Donc la planification présente la phase de définition des moyens (quand et comment ?). Dans le *Dixeco de l'entreprise* (1991), la planification consiste à définir des projets d'action coordonnés dans le temps (période pluriannuelle). Le plan apparaît comme une procédure volontariste (car il

s'agit d'un acte ferme) et rationnelle (car il s'appuie sur des procédures formelles et sur une vision claire et arrêtée de l'avenir). Ce qui fait que la planification se différencie de la démarche adaptative et du pilotage intuitif. Cependant il y a des plans qualifiés de plans flexibles ou glissants qui sont révisés au fil du temps et qui s'adaptent aux changements et aux turbulences de l'environnement. La planification a pour mission d'assurer l'intégration des différentes parties d'une organisation (services et fonctions) et notamment la mobilisation du personnel par rapport aux buts et objectifs poursuivis.

À partir de ces définitions extraites de quelques dictionnaires de référence en « sciences de gestion », nous pouvons déduire la structure conceptuelle des trois concepts fondamentaux en prospective. La structure conceptuelle s'articule autour de trois points : le noyau commun, les différences spécifiques, les nuances (c'est-à-dire les relations entre les termes).

La prospective, la prévision et la planification sont des formes différentes d'anticipation. Toutes les trois suivent une démarche scientifique, anticipative et normative. Leur objet d'étude demeure le futur, l'incertitude, les phénomènes aléatoires et la prise de risque.

- Pour la prospective, le futur est multiple. On parle de différents scénarios (appelés des scénarios possibles) en se basant sur des faits porteurs d'avenir et des signes de rupture. Les acteurs sont actifs (jeu d'acteurs) et construisent le futur voulu et attendu (la liberté d'action et le volontarisme) en ayant une vision large. La démarche prospective est basée sur l'anticipation des ruptures. Son rôle est de comprendre les mécanismes afin d'agir dans le présent.
- La prévision, quant à elle, considère que le futur est unique. On parle de tendances lourdes qui présentent une continuité du passé. Il suffit alors de se baser sur des données quantitatives passées pour prévoir un événement futur (il s'agit d'un événement prévisible et inéluctable). Les acteurs sont alors passifs. L'avenir est subi et inévitable. La vision du futur demeure étroite. La démarche de la prévision est basée sur l'extrapolation des tendances lourdes. Son rôle est de déterminer un événement futur à partir de certaines données quantitatives passées.
- La planification considère également que le futur est unique. Elle se base sur des données aussi bien quantitatives que qualitatives et utilise les résultats prévisionnels. Les acteurs se caractérisent par le volontarisme et la détermination. Ils ont des objectifs et des plans d'action à réaliser. L'avenir est voulu mais la vision du futur est étroite. Son rôle est de définir ce qu'on veut obtenir en préparant un plan d'action et en se basant sur les données prévisionnelles. La prospective est la forme moderne de l'anticipation. Elle est née à la suite de l'échec de la prévision (ses erreurs répétées). Même si elle a recours à différents scénarios possibles, elle utilise des données prévisionnelles. La prévision demeure, malgré sa crise, une forme d'anticipation fondamentale qui complète toutes les autres. La planification, liée aux deux premières, est considérée comme la phase finale de chaque démarche anticipative, la phase d'exécution.

En conclusion, on peut dire que, même si ces trois formes d'anticipation semblent différentes au départ, elles demeurent convergentes et surtout complémentaires.

2. CHOIX DU *CORPUS* :

Un *corpus* est un ensemble fini d'énoncés pris pour objets d'analyse. Plus particulièrement, un ensemble fini d'énoncés considérés comme caractéristiques du type de langue à étudier, réunis pour servir de base à la description et, éventuellement, à l'élaboration d'un modèle explicatif de cette langue. Selon la recherche envisagée, il s'agit donc d'une collection de documents, soit oraux (enregistrés et/ou transcrits), soit écrits, soit oraux et écrits.

Les dimensions du *corpus* varient avec l'objectif du chercheur et la masse des énoncés considérés comme caractéristiques du phénomène à étudier. Un *corpus* est dit exhaustif quand il comprend tous les énoncés caractéristiques. Il est dit sélectif quand il n'en comprend qu'une partie (*Le Dictionnaire de didactique des langues*, 1976).

Nous retenons ici les caractéristiques fondamentales du *corpus*. Un *corpus* doit être homogène, exhaustif et représentatif.

2.1. L'homogénéité :

Elle se définit par :

- la structure énonciative : qui dépend de l'émetteur et du récepteur.
 - Si l'émetteur est spécialiste et le récepteur est spécialiste alors le discours est spécialisé.
 - Si l'émetteur est spécialiste et le récepteur est non identifié, alors le discours est semi-spécialisé et hétérogène.
 - Si l'émetteur est spécialiste et le récepteur est non-spécialiste alors le discours est général.
- l'unité thématique : elle se définit par la cohérence terminologique, par l'invariance de l'objet et par l'homogénéité des champs notionnels.
- l'idiosyncrasie linguistique : on peut citer à ce titre l'exemple du discours académique qui est affirmatif et se caractérise par un nombre réduit de modalisateurs.

2.2. L'exhaustivité :

Elle est assurée par :

- • la diversité interne (le même objet est étudié sous différents angles),
- • l'aspect contrastif (mise en contraste de deux éléments constitutifs de l'objet d'une façon interne ou externe)
- • l'approche modulaire (l'étude approfondie d'un composant du *corpus*)

2.3. La représentativité :

La représentativité, quant à elle, présente la synthèse et le croisement des deux premières dimensions. Elle est conçue dans un cadre prototypique (un prototype peut être contredit mais jamais remis en cause). Le *corpus* peut servir de prototype.

3. L'ANALYSE :

En tenant compte de ces considérations épistémologiques, nous exposons trois échantillons différents de discours : le discours académique, le discours des praticiens et le discours général.

3.1. Le discours académique :

Il s'agit d'un extrait tiré du *Traité élémentaire de prévision et de prospective* d'André-Clément Découflé (1978 : 102). Il est structuré autour de trois concepts fondamentaux : prévision, décision et action. En dégageant les éléments définitoires de ces concepts figurant dans le texte, on distingue :

– des définitions directes :

- Action : système de décisions appliquées : « Les relations entre prévision et action – celle-ci étant prise ici au sens de système de décisions appliquées – ont longtemps été simples. Elles ne le sont plus. »
- Événement : compris comme un hasard non maîtrisé : « Tout prévoir, de sorte que l'événement, compris ici comme un hasard non maîtrisé, ne décide pas à la place de l'acteur. »

– des définitions indirectes : le terme de prévision est défini indirectement et dépend de plusieurs variables et de ses attributs (acteurs, stratégies, objectifs et fins) : « Dans un système d'action où les acteurs sont peu nombreux et aisément identifiables, les stratégies clairement définies, les objectifs précis, les fins dernières dénuées d'ambiguïté, la prévision entretient des rapports directs avec l'action et contribue à réduire effectivement la part du jeu des facteurs aléatoires. »

Donc nous constatons que ce texte se caractérise par une consistance terminologique. Tous les concepts ont des contreparties bien délimitées et bien définies. Pour ce faire, l'auteur a procédé de deux manières différentes : d'une manière directe (en utilisant des définitions simples) et d'une manière indirecte (en utilisant des définitions par la négative). Dans le premier cas, on a une relation d'équivalence entre le terme et sa définition ; dans le second,

il suffit de restructurer toutes les données fournies dans le contexte pour reconstruire la définition.

3.2. Le discours des praticiens, le rapport annuel d'AXA (2004) :

Dans ce rapport nous constatons que l'unité lexicale prospective a été utilisée dans un sens large et flou. On parle de « dimension prospective à moyen terme », d'« approche prospective » et de « base prospective ». Le mot est flou et se limite au rôle d'adjectif : « prospective » qui correspond au futur et à l'avenir ou à une démarche anticipatrice. Par contre, le concept de prévision est employé d'une manière beaucoup plus précise. On parle de « prévision consolidée qui constitue le budget du Groupe et qui se décline en objectifs ». Le concept est un terme qui englobe la démarche et la discipline de prévision. Nous citons

quelques exemples :

[1] Ils examinent périodiquement les analyses et *prévisions* à trois ans. Ces *prévisions*, qui prennent en compte des scénarios extrêmes de variations des marchés financiers, sont également mises à jour mensuellement dans un tableau de bord destiné au Directoire (p. 34).

[2] L'exercice de planification stratégique d'AXA a pour but d'exercer un contrôle en amont sur les orientations majeures et les *prévisions* à trois ans des principales sociétés du Groupe. Après différentes phases d'analyse et d'ajustements, il aboutit à une *prévision* consolidée qui constitue le budget du Groupe et qui se décline en objectifs (p. 29).

[3] La DCFG effectue en continu une *prévision* à 3 ans, en utilisant des scénarios extrêmes sur l'évolution des marchés des actions et des taux d'intérêts (p. 35).

[4] Certaines déclarations figurant ci-après donnent des perspectives et contiennent des *prévisions* qui portent notamment sur des événements futurs, des tendances, projets ou objectifs (p. 138).

[5] [...] des objectifs chiffrés (chiffre d'affaire, dépenses, indicateurs de rentabilité, de productivité et de qualité) basés sur un scénario central de *prévisions* économiques (p. 29). Par ailleurs, dans certains passages du rapport annuel, cette même unité lexicale *prévision* est floue et synonyme du mot « estimation » qui peut correspondre à un chiffre bien déterminé. Nous constatons également la présence de valeurs et de chiffres précis. Quelques exemples suffisent pour illustrer cet emploi :

[1] Axa a publié ses résultats annuels pour l'exercice 2004 le 24 février 2005. Le résultat net part du Groupe, qui s'inscrit à 2 519 millions d'euros, était supérieur aux *prévisions* des analystes et en hausse de 151 % par rapport à 2003 (p. 46).

[2] l'économie reste soutenue par le nombre croissant de visiteurs venus du continent chinois, qui devrait selon les *prévisions* atteindre 20 millions en 2005, après 12 millions en 2004 (p. 139).

[3] La valeur de réalisation est estimée à partir des informations de marchés disponibles ou des *prévisions* de cash-flows attendus, compte tenu de l'intégration d'une prime de risque (p. 225).

Dans ces exemples, il est clair que pour « prévision », il s'agit d'un mot de la langue générale.

3.3. Le discours général :

Pour analyser ce type de discours général, nous nous sommes basée sur le message du président de L'Oréal extrait du rapport annuel 2004. Ce message s'inscrit dans une perspective d'avenir. Il utilise les variables dans une configuration compliquée (aucun terme n'est précis). Cela montre que le discours est écrit par des spécialistes et s'adresse au grand public (récepteur non spécialiste). La preuve, c'est l'absence de consistance conceptuelle ayant pour contrepartie un maillage terminologique serré. Le discours général, même s'il comporte l'essentiel des notions liées aux concepts, ne peut en aucune façon prétendre au statut du discours spécialisé parce que les concepts y sont dissous.

L'étude de ce texte nous a permis de faire les constats suivants :

- aucune terminologie n'est employée,
- tous les ingrédients des trois items y figurent (prospective, prévision et planification) et
- une parfaite dissolution de ces concepts dans le discours sans contrepartie terminologique.

Arrivée à ce point d'analyse, nous pouvons dire que notre hypothèse est globalement vérifiée. Nous avons pu vérifier que les deux extrémités ne posent pas de problèmes : le discours spécialisé se caractérise par sa consistance terminologique et le discours général, dans lequel le mot est roi, obéit aux règles générales. Ce qui est intéressant est d'étudier le discours hétérogène où l'unité lexicale oscille entre les statuts de terme et mot.

4. CONCLUSION

En guise de conclusion, nous pouvons dire que notre analyse peut être exploitée pour mesurer le degré de spécialisation d'un discours. Il serait intéressant de concevoir un programme fondé sur des corrélations entre terme et structure conceptuelle correspondante. Pour chaque domaine, il suffit de dégager la terminologie la plus représentative, d'en préciser le contenu avec consistance et de l'appliquer à des *corpus* variés. Un tel outil serait d'une grande utilité pour l'extraction des données sur le web et aiderait sûrement à faire une typologie de discours empiriquement mesurable.

BIBLIOGRAPHIE

- Baumard (P.), 1996 : *Prospective à l'usage du manager*, « Les essentiels de la gestion », Litec.
 Berger (G.), 1958 : « L'attitude prospective », dans *Prospective*, n° 1.

- Boutinet (J.-P.), 1999a : « Anthropologie du projet », 5^e édition, *Psychologie d'aujourd'hui*, PUF, Paris.
- Boutinet (J.-P.), 1999b : *Psychologie des conduites à projet*, « Que sais-je ? », 3^e édition, PUF, Paris.
- Burland (A.), Eglem (J.-Y.), Mykita (P.), 1995 : *Dictionnaire de gestion*, Foucher.
- Cohen (E.), 1994 : *Dictionnaire de gestion*, La Découverte.
- Condor (R.), 2002, *La conception de projets dans les petites entreprises : une approche par les conduites d'anticipation*, thèse de doctorat de l'IAE de Caen en sciences de gestion. Decoufle (A.-Cl.), 1972 : *La prospective*, Presses universitaires de France, « Que sais-je ? »
- Decoufle (A.-C.), dir., 1978 : *Traité élémentaire de prévision et de prospective*, PUF.
- Dixeco de l'entreprise : *pour comprendre le langage de l'entreprise*, Dunod, 1991.
- Dubois (J.), Giacomo (M.), Guespin (L.), Marcellesi (C. et J.-B.), Mevel (J.-P.), 1973 : *Dictionnaire de linguistique*, Larousse.
- Galisson (R.), Coste (D.), 1976 : *Dictionnaire de didactique des langues*, Hachette.
- Godet (M.), 1992 : *De l'anticipation à l'action*, Dunod, Paris.
- Godet (M.), 1977, *Crise de la prévision, essor de la prospective*, PUF.
- Jouvenel (H.) De, 1999 : « La démarche prospective : un bref guide méthodologique », dans *Futuribles*, n° 247.
- Le Duff (R.), 1999 : *Encyclopédie de la gestion et du management*, édit., Dalloz.
- Lerat (P.), 1995 : *Les langues spécialisées*, PUF.
- Martinet (A.-C.) et Silem (A.), 1989-2003 : *Lexique de gestion*, Dalloz, Paris, (2^e édition).
- Mahe de Boislandelle (H.), 1998 : *Dictionnaire de gestion : vocabulaire, concepts et outils*, Economica.
- Mangueneau (D.), 1976 : *Initiation aux méthodes de l'analyse du discours*, Hachette université.
- Mejri (S.), 2003 : *L'Anticipation entre précision conceptuelle et pratique d'entreprise : cas des entreprises des pays émergents*, DEA en Sciences de gestion à l'Institut d'Administration des Entreprises de Basse Normandie.
- www.axa.fr
- www.loréal.fr

PRÉPONDÉRANCE DU CONTEXTE EXTRALINGUISTIQUE DANS LA CONSTRUCTION DU SENS : L'EXEMPLE DES COMMUNICATIONS DE TRAVAIL DANS LA NAVIGATION AÉRIENNE

Pascale Vergely

Université de Toulouse-Le Mirail

laboratoire ERSS (équipe de recherche en syntaxe et sémantique)

Toulouse, France

vergely@univ-tlse2.fr

***Résumé :** nous envisageons sous un angle nouveau la problématique du contexte extralinguistique pour l'analyse des communications de travail. Notre objectif est de montrer le rôle prépondérant du contexte extralinguistique dans la détermination du sens (et de son interprétation) dans une des situations de travail d'urgence : celle de la navigation aérienne. Au-delà des critères linguistiques définissant l'existence des langages opératifs, nous montrerons que les notions extralinguistiques d'expert d'une part, et de domaine d'autre part, participent activement à l'établissement (ou non) d'un sens commun entre les locuteurs et donc plus largement, au fonctionnement des langages opératifs et à leur variation.*

***Mots-clés :** linguistique de corpus, langage opératif, expert, domaine*

1. INTRODUCTION

Cet article envisage sous un angle nouveau la problématique du contexte extralinguistique pour l'analyse des communications de travail. Il se donne pour objet de montrer le rôle prépondérant du contexte extralinguistique dans la détermination du sens (et de son interprétation) dans une des situations de travail d'urgence. Le domaine concerné est celui de la navigation aérienne. Le *corpus* porte sur des communications téléphoniques entre deux populations d'experts au travail qui communiquent lors d'un dysfonctionnement technique d'un des outils de contrôle utilisé par les aiguilleurs du ciel. Comme le mentionne J. Boutet (1989), le choix des mots est un composant du sens des phrases. Mais nous nous accordons avec l'auteur pour dire que s'il est nécessaire, il n'est pas suffisant. Le sens de l'énoncé implique particulièrement dans ces dialogues opératifs d'urgence, la référence au contexte extralinguistique. C'est dans cette optique que nous situons les propos qui vont suivre.

Au-delà des critères linguistiques définissant l'existence des langages opératifs (Falzon 1989), nous montrerons que les notions extralinguistiques d'expert d'une part, et de domaine (contexte d'énonciation) d'autre part, participent activement à l'établissement (ou non) d'un sens commun entre les locuteurs et donc plus largement, au fonctionnement de ces langages et à leur variation. Parce que le langage est le support des opérations socio-

cognitives (raisonnement, coconstruction avec autrui...), il est communément admis que ce langage d'action ne peut pas être étudié isolément, tant du point de vue des disciplines qui l'abordent que du point de vue de son contexte de production.

2. CONTEXTE DE L'ÉTUDE

L'étude concerne une analyse linguistique réalisée au Centre d'études de la navigation aérienne¹ (CENA). Les dialogues étudiés sont issus de communications téléphoniques entre deux populations d'opérateurs experts au travail du domaine de la navigation aérienne. Ils interviennent dans le cadre d'une mission commune de restitution de la disponibilité des outils utilisés par les contrôleurs. Il s'agit de résoudre les problèmes que les dysfonctionnements techniques entraînent sur les positions de contrôle. Dans cette situation d'urgence, les énoncés du dysfonctionnement technique doivent se faire le plus rapidement mais surtout le plus précisément possible pour que le rétablissement puisse avoir lieu.

3. LES LANGAGES OPÉRATIFS : DÉFINITION

Dans notre étude, c'est précisément sur les langages/dialogues opératifs que nous nous focalisons. Une des principales particularités de ces dialogues réside dans leur visée téléologique : ils sont orientés vers une tâche, un but. Mais ce qui les distingue des dialogues finalisés par exemple, c'est que les dialogues opératifs convoquent exclusivement des locuteurs dits experts dans le domaine de connaissance sur lequel portent leurs propos. Ainsi, selon Falzon² (1989, 1996) toute pratique collective, aussi bien dans les hobbies que dans les activités de travail : « génère ainsi un langage particulier, opératif, qui possède un vocabulaire, une syntaxe, une sémantique spécifiques. [...] Ces langages opératifs présentent plusieurs caractéristiques : ils ne sont utilisables que dans le cadre de la tâche qui a donné lieu à leur élaboration et qu'entre spécialistes du domaine, sous peine de perdre leur valeur communicative, et ils perdent leur efficacité si la situation est inhabituelle » (1996 : 4).

Ces spécificités ainsi délimitées, nous présentons ci-dessous les principaux critères (III) syntaxiques, lexico-sémantiques et pragmatiques qui constituent les langages opératifs. Nous argumentons ensuite (IV) en faveur de l'intégration des notions extralinguistiques d'expert et de domaine pour définir les langages opératifs.

3.1. Du point de vue syntaxique

La plupart des travaux sur les langages spécialisés ou opératifs (Kocourek, 1991 ; Kittredge, 1982 ; Harris *et al.*, 1989 ; Cabré, 1998 ; etc.) s'accordent pour définir ces

¹ Les résultats que nous présentons sont issus d'un travail de thèse (Vergely, 2004) prenant place dans le cadre d'une collaboration entre un laboratoire de recherche en linguistique d'une part, l'Équipe de recherche en syntaxe et sémantique (ERSS) et un organisme d'autre part, le Centre d'études de la navigation aérienne (CENA). L'objectif était d'élaborer une grammaire dans les énoncés de *l'expression du dysfonctionnement technique*.

² Précisons dès à présent que les travaux de Falzon intègrent également la dimension pragmatique (au sens large), ce que ne font pas, par exemple, des travaux plus linguistiques sur les sous-langages.

derniers comme usant de règles syntaxiques réduites *i.e.* d'une grammaire différente de la langue générale. En prenant le cas des communications contrôleurs-pilotes de la navigation aérienne, Falzon (1996) fournit une illustration particulièrement nette de cette syntaxe restreinte. Concernant à proprement parler les langages opératifs, l'auteur (1986 : 63 ; 1989 : 70) expose les cinq caractéristiques syntaxiques de ces langages :

- Certaines règles syntaxiques sont privilégiées ;
- Elles sont isolables par différentes méthodes, qui permettent de construire la grammaire du langage opératif considéré ;
- Ces grammaires sont plus restreintes que la grammaire générale de la langue ;
- Ces grammaires ne sont pas un sous-ensemble de la grammaire de la langue
- Ces grammaires varient d'un langage opératif à un autre.

Il soumet des exemples d'études qui confortent chacun des cinq points. Mais il est important de retenir que ces critères sont établis à partir de l'analyse de situations de communication contrôleurs/pilotes qui utilisent un langage contrôlé, une phraséologie. Cependant, certaines propriétés des langages opératifs ne peuvent pas servir, selon nous, pour rendre compte du fonctionnement général de tout langage opératif. Le langage contrôlé n'est pas le type de langage représentatif des situations de travail, bien au contraire. Il faut reconnaître qu'hormis dans le contexte de la navigation aérienne (civile ou militaire), certaines situations hospitalières (dans les blocs opératoires, etc.), ou dans le nucléaire, l'emploi récurrent d'un langage opératif contrôlé n'est pas courant ou du moins permanent (comme par exemple certaines annonces à la SNCF). C'est tout au moins ce que nous défendons, du point de vue syntaxique (les aspects lexicaux, sémantiques et pragmatiques sont en revanche plus caractéristiques), pour les dialogues chef de salle/superviseurs de la navigation aérienne que nous avons étudiés.

Les résultats de notre étude (Vergely 2004) révèlent l'usage de règles syntaxiques similaires à celles décrites dans des travaux portant sur le langage général. Parce qu'il entre en usage dans un domaine de travail entre experts et parce qu'il est le produit d'une pratique (Falzon 1989), ce langage est donc bien opératif. Mais il s'agit d'un langage opératif non contrôlé dans la mesure où les opérateurs de ces dialogues ne s'expriment pas au travers d'un langage restreint mais au travers du langage naturel. Cette précision est selon nous déterminante et est à la base des caractéristiques des langages opératifs.

3.2. Du point de vue lexico-sémantique

Le critère lexico-sémantique et en particulier la fréquence du vocabulaire technique est certainement l'aspect le plus récurrent des études sur les langages dits de spécialité, ce qui confère à ces études des directions uniquement lexicales. En effet, c'est principalement l'usage du vocabulaire qui permet de faire contraster une langue spécialisée avec la langue établie comme celle de référence, à savoir la langue commune. Pour Falzon (1989 : 66) les termes des langages opératifs sont en nombre « restreint » par rapport à la langue générale. Ce lexique peut « comporter des mots rares pour la langue générale [...] les mots qu'ils [les lexiques] comportent peuvent lui être totalement spécifiques ». En revanche, dès lors que

L'on aborde l'aspect sémantique de ce type de lexique, le fonctionnement récurrent cité dans quasiment toutes les études est celui de la monosémie. Ce phénomène sémantique reste un fonctionnement récurrent cité dans quasiment toutes les études, Falzon (1989 : 75) n'en fait pas l'exception, puisque selon l'auteur : « les lexiques des langages spécialisés peuvent comporter des mots spécifiques à ces langages : ces mots seront monosémiques » Cabré (1998) va également dans ce sens en expliquant que ces langages sont « univoques ».

L'exemple suivant issu de notre *corpus* révèle que cette affirmation sur la monosémie est contestable dans le langage opératif étudié.

L'exemple que nous prenons, et plutôt le contre-exemple, est le terme *radar*. Ce terme, bien que de nature non ambiguë aujourd'hui dans ces communications de travail, recouvre deux significations (et plutôt deux sens puisque nous sommes en discours) suivant le type de locuteur. Il renvoie, pour une population d'opérateurs, à l'antenne (radar) située à l'extérieur des centres de contrôle, alors que pour l'autre population il désigne le scope ou l'image radar utilisée par les contrôleurs aériens. D'autres exemples issus d'une étude (Vergely 2002) réalisée pour le CENA mettent au jour des facteurs d'incompréhensions et d'ambiguïtés gênant le déroulement du dialogue. Mais cet exemple suffit cependant pour démontrer que la monosémie du lexique ne peut pas être généralisable à tout langage opératif. Elle est au moins invalidée dans notre *corpus* d'étude.

3.3. Du point de vue pragmatique

Les trois traits linguistiques précédents sont couramment considérés comme critères définitoires absolus des textes et langages spécialisés. Ce n'est que plus rarement que la dimension pragmatique est introduite dans ces définitions. Cette dimension s'attache spécialement à l'interprétation de ce qui est dit, ce qui revient à prendre en compte non seulement le locuteur mais aussi l'interlocuteur. Les travaux de Cabré (1998 : 138) proposent l'intégration de ces deux situations (du locuteur et de l'interlocuteur) en introduisant les aspects pragmatiques *i.e.* « le sujet, les utilisateurs et les situations de communication » et les aspects fonctionnels, c'est-à-dire les objectifs visés, dans sa définition d'une langue de spécialité. Autrement dit, il est important de souligner que l'interprétation des dialogues orientés par un objectif spécifique se situe à l'intersection de données linguistiques et de données situationnelles (contexte, interlocuteurs...), ces dernières participant incontestablement à la définition de ces dialogues. Mais parce que les mécanismes d'interprétation dépendent le plus souvent de critères extralinguistiques spécifiques à chacun des domaines de travail, il nous semble malaisé de vouloir généraliser l'ensemble des critères définitoires (syntaxique, lexico-sémantique et pragmatique) des langages opératifs à toutes les situations de dialogues de travail entre experts (Condamines et Vergely 2005). L'argument fourni selon lequel le recours au langage opératif correspond au fait que les interlocuteurs ont en commun non seulement un objectif, mais aussi un ensemble de connaissances n'est pas, selon nous, un trait de caractérisation suffisant. Partant de ce constat, nous proposons d'une part une définition alternative des langages opératifs qui tienne compte du fait que certaines situations de travail font appel au langage naturel et plus uniquement au langage contrôlé. D'autre part, nous intégrons deux autres éléments à la définition des langages opératifs : la notion d'*expert*, notion hétérogène étant

donné le degré de spécialisation au sein du groupe de locuteurs et la notion de *domaine* et plus précisément de domaine d'activité différent.

4. POINT DE VUE EXTRALINGUISTIQUE : IMPORTANCE DES NOTIONS D'EXPERTS ET DE DOMAINE

Si les traits linguistiques sont des critères déterminants pour définir l'existence des langages opératifs, il nous semble maintenant évident que les deux notions extralinguistiques d'experts et de domaine participent activement au fonctionnement des langages opératifs et à leur variation. Bien que toujours mentionnées dans les études qui traitent des spécificités des communications de travail, les définitions d'expert et de domaine sont souvent, selon nous, présentées de manière superficielle. Le plus couramment, un expert est une personne spécialiste d'un domaine donné et qui dispose de fait, de connaissances spécifiques à ce domaine³. Le domaine correspond à l'ensemble des circonstances extralinguistiques dans lesquelles s'insèrent les productions verbales et plus largement les actions (langagières ou non) étudiées. Dans notre cas, le domaine est celui de la navigation aérienne. Les parties à suivre montrent comment ces deux notions mettent en œuvre dans un cas, des distinctions du point de vue des connaissances (experts), et dans l'autre, des distinctions du point de vue de l'interprétation (domaine).

4.1. Distinction du point de vue de connaissances : la notion d'expert

L'analyse descriptive des situations de consultation médicale a permis à Lacoste (2001) de mettre en évidence certains faits saillants et récurrents dans les comportements communicatifs des participants. L'un d'entre eux concerne les savoirs inégaux mis en jeu dans ce type de consultation dus au fait que sont en présence un expert et un novice. Ce constat donne lieu à ce que Lacoste nomme avec pertinence « l'asymétrie des rôles » et qui est, toujours selon l'auteur, « définitoire de la situation ». Nous allons montrer que cette notion d'asymétrie est loin d'être exclusivement réservée aux communications entre experts et non-experts mais peut être étendue à certaines situations de communications entre experts. Les explications de cette asymétrie trouvent leurs origines dans la formation et les objectifs différents entre les deux groupes d'opérateurs.

Les deux types de locuteurs partagent effectivement des connaissances communes liées au domaine de travail dans lequel ils interagissent : celui de la navigation aérienne et plus spécifiquement, le domaine relatif au maintien de la disponibilité des moyens de contrôle. Mais ces connaissances sont en fait également asymétriques dans la mesure où une population de locuteurs met en œuvre de par leur métier et leur formation de base, un degré de perception distinct de l'autre population.

Les deux exemples ci-dessous fournissent les conséquences de cette distinction sur le dialogue.

³ Si la notion de domaine se limite, dans notre cas, à l'environnement de travail, elle n'en est pas en revanche exclusive. Le domaine pourrait aussi bien renvoyer à celui du sport (spéléologie, voile, etc.), du jeu (cartes, échecs, etc.)...

Le premier exemple concerne la prise en compte du dysfonctionnement d'une fréquence radio utilisée par les contrôleurs pour entrer en contact avec un aéronef. Celle-ci varie en fonction du type d'opérateur. Pour l'un (le chef de salle), la priorité se pose en termes d'exploitation: il s'agit de savoir si la présence du dysfonctionnement permet tout de même d'exploiter la fréquence radio en cause. Dans le cas d'une réponse négative, la détermination d'un délai de temps concernant l'inexploitation est nécessaire. Cette problématique est prépondérante dans la mesure où elle lui permet d'évaluer les éventuelles répercussions au niveau de la salle de contrôle (changement ou non de fréquence par exemple). Pour l'autre (le superviseur technique), la gestion de l'événement est différente. Il s'agit en premier lieu d'évaluer le degré de gravité du problème. L'expression du dysfonctionnement technique doit donc être des plus précises pour pouvoir identifier quel élément technique (et il peut y en avoir plusieurs) est réellement à la source du problème. En somme, ce qui est de l'ordre de l'information pertinente pour une population d'experts correspond souvent à une information générale (et imprécise) pour la seconde population d'experts. À ceci s'ajoute le fait qu'il est parfois difficile, pour le chef de salle, d'exprimer techniquement le dysfonctionnement auquel il a affaire. Son métier, sa formation, ses objectifs de travail sont des arguments qui expliquent la représentation générale dont dispose le chef de salle concernant le fonctionnement des systèmes techniques qu'il utilise. Corrélativement, le métier, la formation et les objectifs de travail des superviseurs techniques déterminent ces derniers comme *experts* du fonctionnement des systèmes techniques exploités par les contrôleurs. Mais la diversité et la complexité des systèmes techniques conduisent les superviseurs techniques à se spécialiser du point de vue de la formation. Un superviseur technique sera ainsi spécialiste de la chaîne radio (fréquence), un autre de la chaîne radar, etc.

Il y a donc plusieurs types d'experts et également des experts spécialistes qui disposent chacun de connaissances et compétences distinctes et spécifiques. Il découle de cette asymétrie des connaissances des problèmes d'ambiguïtés comme l'illustre ce second exemple: les formations et les métiers différents des opérateurs entraînent certaines incompréhensions dans les dialogues. Dans l'extrait suivant :

[1] CDQ : –Ouais heu j'te demandais est-ce qu'il serait possible ce matin de faucher le au niveau du 33 heu droite ?

[2] CDT : –Alors qu'est-ce que tu veux faucher au niveau du 33 droite ? qu'est-ce t'appelles le 33 droite ? le seuil ?

[3] CDQ : –Le, le loc au niveau du loc.

[4] CDT : –Ha le loc 33 droite.

[5] CDQ : –Ouais.

[6] CDT : –Déjà c'est pas la même chose.

[7] (...)

La prise de parole [2] matérialise explicitement le fait que le contrôleur veut s'assurer de bien interpréter la demande d'intervention du MO et plus précisément son lieu d'intervention. Bien que le réajustement définitoire s'opère de manière rapide, il n'en reste

pas moins qu'il est nécessaire. Celui-ci se justifie à la prise de parole [6] du CDT : « c'est pas la même chose ». Les priorités de travail (techniques et opérationnelles) et le domaine d'expériences sont des facteurs qui modèrent l'accès à un référentiel commun. La prise de conscience de ces facteurs se matérialise par des réajustements de connaissances : ici, la « 33 droite » réfère par défaut au seuil pour un contrôleur (niveau du toucher des roues) alors qu'il renvoie en revanche, par défaut, pour la MO au *localizer* situé à l'opposé de la piste (seuil 15 gauche).

L'usage de termes spécialisés entre les opérateurs doit être partagé par les deux populations d'opérateurs pour que la communication soit compréhensible et donc plus efficace. Il semble nécessaire que les opérateurs mesurent la part non négligeable de l'implicite sous-jacent à l'utilisation de termes propres à chacun de ces univers de connaissances (et de métiers). Le choix d'un terme (pour le locuteur), comme celui de son interprétation (pour l'allocutaire), est fondamentalement une question de « point de vue » (Condamines et Rebeyrolle 1996, 1997 ; Condamines 2003 : 98). Cette notion est selon nous centrale dans la définition des langages opératifs. Dans cette optique et à la lumière des remarques soulevées jusqu'ici, la question du domaine inhibe celle de point de vue. Et c'est bien cette notion identifiée par les auteurs de point de vue et de points de vue différents d'un même objet (ou concept) qui donne lieu à la diversité terminologique, à la complexité et à l'ambiguïté des échanges mais aussi incontestablement, à la réhabilitation de la variation. Pour ce dernier point, le contexte extra-linguistique prend une place centrale. Il est, comme le note Condamines (2003 : 17), « un risque pour la stabilité ». Cette observation nous conduit à aborder le second aspect : celui du domaine. Celui-ci doit s'entendre en tant qu'espace ouvert (Gaudin 1993) imbriquant différents champs d'activités et différents experts spécialistes.

4.2. Distinction du point de vue de l'interprétation : la notion de domaine

L'importance de la notion de domaine tient à l'étroite relation qu'elle entretient avec celle de communication. Il y a, comme le note Boutet (2003), interdépendance du langage et du contexte d'action : le caractère fonctionnel ou finalisé des communications de travail en est une face visible. Les communications opératives (et plus largement toute communication de travail) s'interprètent de façon particulière en fonction du domaine dans lequel elles s'insèrent. C'est le domaine qui actualise le sens, voire parfois la signification et l'interprétation des communications. Comme le mentionne Grosjean et Lacoste (1999 : 13), « un énoncé n'acquiert véritablement son sens qu'en contexte ». L'exemple du participe passé « hachée » en est une illustration. Son sens renvoie, par défaut dans le *corpus* d'étude de la navigation aérienne, à un dysfonctionnement technique alors que cette valeur de dysfonctionnement ne se vérifie pas dans le domaine culinaire par exemple. C'est donc bien le domaine d'énonciation qui élabore le sens des objets de discours. Autrement dit, le contexte extra-linguistique « influence le sens que l'interlocuteur va donner à un énoncé » (Condamines 2003 : 19). Dans notre cas, c'est le domaine (communicationnel) qui fait qu'on parle de langage opératif. Que l'on se place du point de vue du linguiste, du sociologue, du sociolinguiste, ou de l'ergonome ne change en rien le fait que la notion de domaine (ou de contexte) soit considérée comme une notion centrale dans l'analyse des

communications de travail. Elle est familière à l'ensemble de ces disciplines, elle y est simplement appréhendée de manière plus ou moins explicite.

Le point de vue que nous défendons est le suivant. Nous considérons qu'il faut dépasser le cadre d'une définition univoque de la notion de domaine renvoyant *stricto sensu*, dans les langages opératifs, au domaine de travail. Nous avons argumenté plus haut en faveur d'une définition plurielle de l'expert, qui rende compte de la présence d'experts spécialistes. Dans cette continuité, nous devons envisager une place identique aux domaines d'activités. Le pluriel de domaines d'activités est significatif dans la mesure où il ne renvoie pas, dans notre étude, à l'activité du contrôle aérien – c'est-à-dire dans ce cas, au domaine de travail qui est identique pour tous les experts qui y interagissent (contrôleurs, chef de salle, pilotes, techniciens, etc.) – mais il est relatif aux activités respectives de chacun de ces mêmes opérateurs. Chacun des opérateurs use de connaissances (terminologie) inhérentes à leur spécialité. L'extrait suivant justifie ce constat.

[1] MO : – On a des problèmes de connexion avec le STIP là donc on est en train de voir avec le CESNAC

[2] CDS : – Un problème de connexion avec le STIP qu'est-ce que ça va entraîner ça ?

[3] MO : – Ben pour le moment heu on n'a pas de communication avec les infos du STIP

[4] CDS : – Ça va poser des problèmes ça il va y avoir un paquet de décorrélations alors (...)

Nous voyons que l'énonciation par le MO de l'expression du dysfonctionnement technique des « problèmes de connexion avec le STIP » n'est pas explicite pour le CDS. Le fait qu'il énonce à son tour « qu'est-ce que ça va entraîner » suggère une part de méconnaissance pour le CFDS concernant les effets de ce type d'événement au niveau du contrôle aérien, c'est-à-dire du point de vue de son activité de travail. L'expression « problème de connexion » n'est rendue explicite que lors de la reformulation du MO (3) c'est-à-dire par « pas de communication avec les infos du STIP ». C'est seulement à ce moment que le CDS fait le lien entre la nature du dysfonctionnement « pas de communication » et les conséquences que cela implique pour lui c'est-à-dire des « décorrélations ». Au-delà des difficultés que rencontrent les opérateurs pour verbaliser, en termes clairs pour l'allocutaire, un dysfonctionnement technique, cet exemple montre l'usage de connaissances non partagées voire partiellement partagées entre ces opérateurs. Cette disparité est au cœur de la plupart des ambiguïtés dans le dialogue. Son origine provient du fait que ces deux groupes d'opérateurs ne relèvent pas du même métier. Ils ont un domaine de travail identique – le maintien de la disponibilité des moyens de contrôle – mais des domaines d'activités distincts scandés par des spécialités et des objectifs différents : pour l'un, il s'agit d'assurer les conditions pour un contrôle optimal sur l'ensemble des vols alors que pour l'autre, il s'agit de maintenir et garantir la continuité de service des systèmes techniques.

5. CONCLUSION

L'analyse menée sur ces dialogues entre experts nous a conduit à introduire dans la définition des langages au travail la prise en compte de l'hétérogénéité des notions

d'experts (autrement dit un degré de spécialisation au sein du groupe de locuteurs) et de domaine (plusieurs sous-domaines relatifs aux activités de chacun). Les difficultés résident toutefois dans la prise en compte, par les locuteurs des dialogues, du fait qu'ils disposent de connaissances non partagées. Cette prise en compte permettrait de pallier, au moins en partie, certaines ambiguïtés linguistiques accentuées dans ce type de communications lié à l'urgence des situations. L'ensemble des points mentionnés dans cet article a entre autres pour objectif d'établir une définition des langages opératifs qui soit plus objective : aux traditionnels critères linguistiques et extra-linguistiques se lient de manière intime les notions d'experts et de domaine. Les résultats de notre étude nous ont permis de comprendre qu'il est souvent prématuré de conclure que les discours professionnels mettent en place un fonctionnement linguistique qui leur est propre *i.e.* des règles qui relèvent du domaine de travail étudié. La plupart des dialogues de travail entre experts d'un même domaine met au contraire en œuvre ce que nous avons appelé un langage opératif non contrôlé *i.e.* entraînant un lexique et des connaissances de métiers liés aux spécificités des domaines d'activité des opérateurs, et qui s'actualise au travers du langage naturel (langue commune).

Enfin, soulignons que la prise en compte des variabilités linguistiques et extra-linguistique est également prépondérante pour l'interprétation du sens par les interlocuteurs du dialogue aussi bien que par l'analyste. En prenant en compte précisément ces distinctions du contexte extra-linguistique c'est-à-dire des experts spécialistes et des domaines d'activités distincts, l'analyste appréhende de manière différente le contenu sémantique des échanges et les problèmes d'ambiguïté et de lexique non partagé entre les locuteurs. Il semble dès lors utile d'accorder une place importante aux données issues d'entreprises. Elles fournissent des éléments prépondérants et novateurs à intégrer dans les analyses des *corpus* textuels.

BIBLIOGRAPHIE

- Boutet (J.), 1989 : « La qualification professionnelle entre langue et discours », *Langages*, n° 93, Larousse, pp. 9-22.
- Boutet (J.), 2005 : « La part langagière du travail : un genre discursif ? », *Actes du colloque 8th International Pragmatics Conference, Toronto 13-18 July 2003*, « L'analyse des actions et des discours en situation de travail », sous la dir. de J-P. Bronckart et L. Filliettaz.
- Cabré (M.-T.), 1998 : *La Terminologie : Théorie, Méthode et Applications*, Presses de l'Université d'Ottawa, Armand Colin.
- Condamines (A.), 2003 : *Sémantique et Corpus spécialisés : constitution de bases de connaissances terminologiques*, Habilitation à diriger les recherches en sciences du langage, Université de Toulouse-Le Mirail.
- Condamines (A.) et Rebeyrolle, (J.), 1996 : « Point de vue en langue spécialisée », *Meta*, n° 42, 1, pp. 174-184.
- Condamines (A.) et Rebeyrolle (J.), 1997 : « Utilisation d'outils dans la constitution de bases de connaissances terminologiques : expérimentation, limites, définition d'une méthodologie », *actes des 1^{res} JST Francil*, pp. 529-535.
- Condamines (A.) et Vergely (P.) 2005 : « L'expression du dysfonctionnement technique dans la navigation aérienne : une approche de linguistique de *corpus* » dans Filliettaz (L.) et Bronckart (J.-P.), édit., 2005 : *L'analyse des actions et des discours en situation de travail. Concepts, Méthodes et Applications*.

- Louvain-la-Neuve, Peeters, collection « Bibliothèque des Cahiers de l'Institut de linguistique de Louvain », pp. 177-197.
- Falzon (P.), 1989 : *Ergonomie cognitive du dialogue*, Presses universitaires de Grenoble.
- Falzon (P.) 1996 : *Travailler par le langage. Performances humaines et techniques*, n° hors-série, pp. 3-8.
- Gaudin (F.), 1993 : *Pour une socioterminologie*, Publications de l'Université de Rouen n° 182.
- Grosjean (M.) et Lacoste (M.), 1999 : *Communication et intelligence collective. Le Travail à l'hôpital*, Paris, Presses universitaires de France, coll. « Le Travail humain ».
- Kittredge (R.) 1982 : « Variation and homogeneity of sublanguages » dans Kittredge (R.) et Lehrberger (J.) édit., *Sublanguage. Studies of language in restricted semantic domains*, Berlin, Walter de Gruyter.
- Kocourek (R.), 1991 : *La Langue française de la technique et de la science. Vers une linguistique de la langue savante*, deuxième édition augmentée, Wiesbaden, Brandstetter.
- Lacoste (M.), 2001 : « Analyse de consultations médicales en présence d'un intermédiaire linguistique non professionnel », *Actes du VIII^e congrès de l'Association pour la recherche interculturelle (ARIC)*, Université de Genève, 24-28 septembre 2001, <http://www.unige.ch/fapse/SSE/groups/aric/Textes/Traverso.pdf>.
- Vergely (P.), 2002 : *Analyse sémantique de communications spécialisées : mise au jour de facteurs gênant le dialogue*, Note technique 02-886, Toulouse, CENA.
- Vergely (P.), 2004 : *Analyse linguistique de l'expression du dysfonctionnement technique : le cas des échanges entre chefs de salle et maintenance opérationnelle dans la navigation aérienne*, thèse de doctorat en sciences du langage. Université Toulouse- Le Mirail, 18 octobre 2004.

LE RÔLE DU CONTEXTE DANS LES LEXIQUES TECHNIQUES : LE CAS DE LA MÉTALLURGIE DE RÉCUPÉRATION EN AFRIQUE DE L'OUEST¹

Anneleen Van der Veken

linguistique africaine

Université libre de Bruxelles – Musée royal de l’Afrique centrale

Bruxelles, Belgique

avdveken@ulb.ac.be

Résumé : cette étude consiste en une approche théorique des mécanismes utilisés dans la constitution d’un lexique qui accompagne une nouvelle technique, le vocabulaire des fondeurs de l’aluminium hausa et zarma au Niger. Non seulement les mécanismes du changement lexical sont pris en compte, mais nous démontrons aussi l’influence du contexte sur le lexique utilisé. Afin d’élaborer un arrière-plan théorique plus général, ce lexique est comparé aux vocabulaires d’autres techniques : la poterie, la soufflerie de verre, les chemins de fer et la marine.

Mots-clés : nouvelle technique, nouveau lexique, mots ou termes ?, secteur « informel », hausa, zarma, autres langues

1. INTRODUCTION

Quand une nouvelle technique apparaît dans une société, les récepteurs de cette technique ont besoin d’un vocabulaire pour son exécution quotidienne. Il s’agit non seulement de dénommer les gestes effectués, mais aussi les produits fabriqués, ainsi que les matières premières et les outils utilisés.

Le point de départ de notre étude est le vocabulaire de la métallurgie de récupération au Niger, en Afrique occidentale. Nous analysons dans une première partie les lexiques hausa et zarma de la fonte de l’aluminium qui ont été récoltés par des enquêtes de terrain. De cette manière, les mots et leur sens sont étudiés dans leur contexte réel – qui se situe dans un milieu urbain mais « informel » – ce qui nous semble un aspect indispensable pour la compréhension de la constitution d’un tel lexique. De plus, il s’agit d’un langage lié à l’oralité qui ne peut donc pas être étudié à partir d’un manuel ou d’un dictionnaire. Ce cas d’étude – d’un vocabulaire associé à une technique récente dont nous connaissons plus ou

¹ La présente étude est supportée par le Fonds des bourses de l’Université libre de Bruxelles. Je tiens à remercier mon directeur de thèse Pr Claire Grégoire ainsi que le Dr Koen Bostoen pour leur soutien dans la préparation de cette article, et Muriel Garsou pour la relecture. Ma reconnaissance va aussi à Dr Hamidou Seydou Hanafiou de l’Université Abdou Moumouni de Niamey, pour son aide dans l’analyse des données zarma.

moins la date d'introduction et le lieu d'origine – nous donne l'opportunité de voir comment se constitue un lexique technique nouveau dans cette communauté africaine.

Afin d'élaborer un arrière-plan théorique plus général, le lexique de la métallurgie de récupération doit être comparé aux vocabulaires d'autres techniques. Nous avons choisi un cas similaire mais beaucoup plus ancien, celui du vocabulaire relatif à la poterie dans les langues bantoues. Pour élargir le domaine géographique de cette étude, nous jetterons ensuite un coup d'œil sur le lexique des souffleurs de verre en Belgique au début du XX^e siècle, une activité pratiquée d'une manière peu industrielle. Nous finirons par la comparaison avec les vocabulaires des chemins de fer et de la marine en France.

La conclusion nous permet d'amplifier la discussion et de traiter quelques aspects plus généraux de la constitution des vocabulaires techniques.

2. LE LEXIQUE DE LA FONTE DE L'ALUMINIUM AU NIGER

Bien qu'il existe des différences considérables entre les vocabulaires utilisés par les fondeurs individuels, la plupart des fondeurs partagent une partie importante du lexique. Dans un premier temps, nous analyserons les lexèmes communs, tandis que les différences seront abordées dans une deuxième partie en tenant compte du contexte sociolinguistique et du contexte socioprofessionnel dans lesquels chaque fondeur se trouve en tant qu'individu.

2.1. Les mécanismes du changement lexical

Pour l'analyse des vocabulaires de la fonte de l'aluminium en hausa et en zarma, nous nous baserons en premier lieu sur la classification du changement lexical qui a été proposée par Geeraerts (1997). L'utilisation d'une telle classification facilite la comparaison du lexique dans les deux langues envisagées, qui ne sont pas apparentées. L'approche de Geeraerts étant une approche purement théorique, nous sommes obligés de renvoyer à d'autres classifications pour l'explication de certains phénomènes retrouvés dans les lexiques étudiés. Les exemples traités dans la présente étude sont choisis dans une liste qui en comporte d'autres.

Dans sa classification, Geeraerts (1997) distingue deux catégories principales de changements : les mécanismes onomasiologiques et les mécanismes sémasiologiques. Les mécanismes onomasiologiques conduisent à des changements par lesquels un certain concept – déjà lexicalisé ou non – est exprimé par un nouvel *item* lexical ou un *item* lexical différent. Une première catégorie est celle de la formation de mots. Les mots appartenant à cette catégorie sont dérivés d'autres lexèmes du vocabulaire « standard », à travers des procédés d'innovation lexicale spécifiques à chaque langue particulière, compte non tenu des mécanismes sémasiologiques qui, eux aussi, se basent sur le vocabulaire existant. Dans ce dernier cas, on peut parler de transferts lexicaux par glissement de sens, ce qui sera illustré plus loin. Deux procédés morphologiques majeurs peuvent être distingués dans la catégorie de la formation de mots. Les exemples en [1] illustrent la dérivation qui se fait le plus souvent à partir des verbes et qui concerne surtout les noms d'agent, d'instrument et de lieu ; la composition est illustrée en [2].

[1] noms déverbatifs

hausa mǎkǎ:mǎ:² « la poignée » déverbatif, nom de lieu, du verbe
kǎ:mǎ : « saisir, attraper, prendre »

zarma ìkó:gǎ: « l'aluminium sec » déverbatif du verbe kó:gú « être sec »

[2] noms composés

hausa kǎ:tǎ:kón k'úsǎ « le cadre inférieur du modèle »

« planche »+ « par terre, en bas » (adv.)

zarma gǎndǎ kǎ:tǎ:kò « le cadre inférieur du modèle »

« en bas, à terre » (adv.) + planche

Une deuxième catégorie dans les mécanismes onomasiologiques est celle des emprunts. Les mots d'origine étrangère sont surtout empruntés au français et à l'anglais ; ils sont le plus souvent adaptés à la phonologie de la langue emprunteuse. Quelques exemples d'emprunts au français sont donnés en [3]. Les emprunts à l'anglais sont surtout retrouvés dans les vocabulaires hausa – probablement parce que ces locuteurs ont beaucoup de contacts avec le hausa du Nigeria qui est très influencé par l'anglais, la langue officielle de ce pays. Ce phénomène est illustré en [4]. Le nombre des emprunts anglais en hausa n'est en revanche pas aussi grand que celui des emprunts au français. Il faut par contre noter que beaucoup de ces emprunts au français et à l'anglais sont connus par la plupart des locuteurs hausa et zarma. Ces mots ne sont probablement pas empruntés directement au français ou à l'anglais, ils peuvent être passés par le répertoire commun du zarma et du hausa, ou par le vocabulaire des fondeurs étrangers qui ont introduit la technique au Niger ; une recherche plus approfondie doit tirer cet aspect au clair. Certains fondeurs zarma ont intégré des emprunts au hausa dans leur lexique, ce qui est illustré en [5], tandis qu'aucun Hausa n'utilise un mot d'origine zarma. Quelques éléments qui peuvent contribuer à l'explication de ce phénomène sont traités quand on parle du contexte.

[3] emprunts au français

hausa	ʃǎrbǎ	« le charbon »
pǎl	« la pile »	< pile
tǎjô/tǎjô	« le tuyau »	< tuyau
mǎrtô	« le marteau »	< marteau
ʃǎpô	« le cadre supérieur du modèle »	< chapeau

² Les mots sont transcrits de manière phonétique selon l'Alphabet international de phonétique.

zarma	sà r b ã	« le charbon »
p î l	« la pile »	< pile
t ÿ j ô	« le tuyau »	< tuyau
m â r t ô	« le marteau »	< marteau
s â p ô	« le cadre supérieur du modèle »	< chapeau

[4] emprunts à l'anglais

hausa	bá: t ð r	« la pile »
w é j à:	« le fil de fer »	< wire
r ó b à:	« le caoutchouc »	< rubber
zarma	bá: t í r ì	« la pile »

[5] emprunts au hausa

zarma	m ð k w á: s í:	« l'outil pour arranger
le sable »	< m á k w á: s í:	
m â k á: m ì:	« la poignée »	< m á k á: m á:

Un dernier procédé onomasiologique est la création de mots tout à fait nouveaux dans la langue. La seule manière de créer de tels mots dans le domaine de la fonte de l'aluminium est l'emploi d'onomatopées. Le résultat n'est cependant pas étranger à la langue, étant donné que l'onomatopée a été construite selon les règles phonologiques et morphologiques des langues concernées. L'exemple [6] nous montre le nom donné à « la batte », qui est basé sur le bruit que le fondeur fait en tapant le sable avec la planche.

[6] onomatopées

hausa	p á p p á k á:	« la batte »
zarma	p á k p á k / p ó k p ó k	« la batte »

Outre les mécanismes onomasiologiques, Geeraerts (1997) distingue des mécanismes sémasiologiques du changement lexical. Comme c'était le cas pour certains mécanismes onomasiologiques, des mots du langage commun sont adaptés, ici sur le plan sémasiologique, pour mieux servir le nouveau contexte de la fonte de l'aluminium. Les différents mécanismes onomasiologiques, tant qu'ils sont classifiés par Geeraerts interagissent souvent dans l'analyse d'un mot particulier du lexique de la fonte de l'aluminium. Pour cette raison, nous avons décidé de suivre sa classification, mais de tenir compte de la distinction faite par Crowley (1994 : 148-152) entre les types de changement sémantique et les éléments cognitifs qui dirigent ces changements. Les mécanismes de spécialisation et de généralisation de Geeraerts font partie de la première catégorie, tandis que ceux de la métonymie et de la métaphore rejoignent plutôt la deuxième catégorie. Il est aussi important de souligner le fait que, pour cette analyse, nous avons décidé de ne pas

partir de la langue en général, mais de prendre comme point de départ le lexique tant qu'il est utilisé dans l'atelier de la fonte. De ce point de vue, les mécanismes peuvent avoir lieu dans le lexique des fondeurs, indépendamment de ce qui se passe dans la langue en général. Un premier mécanisme sémasiologique est celui de la spécialisation : des mots avec un sens plus général, qui peuvent être utilisés dans plusieurs contextes, ont une signification très spécifique dans le contexte de la fonte de l'aluminium. Bien que les locuteurs hausa et zarma non spécialisés connaissent le sens des mots concernés, ils ne les mettent pas en rapport avec la technique de la fonte de l'aluminium. Les exemples de [7] illustrent ces mécanismes de spécialisation.

[7] la spécialisation

hausa ɪ̀àŋgà ɪ̀áŋgà « la bande métallique qui tient ensemble le modèle »

signification originale : « personne ou chose longue, mince »

zarma ɪ̀àŋgà ɪ̀áŋgà « la bande métallique qui tient ensemble le modèle »

signification originale : « grand et mince »³

Pour ce qui est du deuxième mécanisme, nous préférons parler d'extension plutôt que de généralisation, comme Geeraerts (1997) le propose. En effet, les termes traités dans cette catégorie sont surtout des termes relatifs à un domaine particulier qui sont récemment entrés dans le lexique de la fonte de l'aluminium, un autre domaine particulier. L'emploi du mot ne s'est donc pas généralisé dans le vocabulaire de tous les locuteurs hausa et zarma, mais reste plutôt restreint. La plupart des mots qui peuvent être classifiés sous la catégorie de l'extension appartiennent au lexique des forgerons. Il faut noter que ni les outils, ni l'emploi de ces outils ne changent beaucoup ; le changement se trouve dans le contexte de l'utilisation de ces outils. Les termes illustrés en [8] sont utilisés par la plupart des fondeurs.

[8] l'extension des termes de la forge

hausa àwà r t ák ì: « la pince »

má g á: g á r í: / d á r t ò: « la lime »

zarma k á m b ù « la pince »

c í r r à « la lime »

Ce ne sont donc pas seulement les termes provenant de la terminologie d'un autre domaine particulier qui sont sujets à ce mécanisme d'extension. La catégorie comporte également des mots d'emploi commun, comme illustré par les exemples de [9]. Il faut noter que l'objet utilisé est toujours le même, bien que l'emploi dans l'atelier ait beaucoup changé. On peut parler d'une extension du centre prototypique qui caractérise les noms concernés. La louche utilisée dans l'atelier de la fonte n'est plus « une cuillère avec laquelle on sert le

³ La traduction du terme en songhay oriental dans la base de données lexicale sur la zone sahel-saharienne de l'Université de Nice, http://sahelia/unice.fr/GDRE_orig.html.

potage », mais elle est devenue « une cuillère avec laquelle on enlève l'excédent d'aluminium », l'aluminium étant comparé au potage de manière métaphorique.

[9] l'extension des mots d'emploi commun

hausa	kò:fí j à:	« le lisseur, pour lisser le sable »	litt. « cuillère »
	l ú d d è j í:	« la louche, pour enlever l'excédent d'aluminium »	litt. « louche »
zarma	k á w r ò	« le lisseur, pour lisser le sable »	litt. « cuillère »
	g ò m b ó	« la louche, pour enlever l'excédent d'aluminium »	litt. « louche »

Les exemples en [10] sont également des illustrations d'extension. Le nom du carreau est employé pour dénommer un morceau de carreau qui est utilisé comme lisseur dans l'atelier. Le processus derrière ce changement sémantique est la métonymie, l'outil est désigné par le matériel dont il est fabriqué.

[10] l'extension par métonymie

hausa	k à r ô	« le lisseur »	< carreau
zarma	k à r ô	« le lisseur »	< carreau

Le dernier mécanisme sémasiologique que nous traitons ici est celui de l'analogie, « the group of analogical changes can be subclassified on the basis of the associative link that exists between the expression that is subject to the semasiological extension and the expression that furnishes the model for the extension. » Geeraerts (1997 : 100). Dans le lexique de la fonte de l'aluminium, la relation entre les deux expressions est toujours de nature syntactique. L'expression qui est sujette à l'extension sémasiologique est formée à l'aide d'ellipse, comme cela est illustré en [11].

[11] l'analogie

hausa	t ú k ú p á: < t ú k ú p á: z í b ì:	« le creuset »	< la marmite à verser
zarma	k ù s ú < d ò: r ì j ó ŋ k ù s ú	« le creuset »	< la marmite à verser

Non seulement le mécanisme de l'analogie joue un rôle dans les exemples en [11], on peut aussi parler de spécialisation. L'objet qu'on dénomme ici n'est pas une marmite comme on l'utilise pour préparer la nourriture ; il s'agit ici d'un creuset – souvent fait d'un compresseur de réfrigérateur – qui est utilisé, comme la marmite, pour chauffer quelque chose sur le feu. Le processus cognitif derrière ce changement lexical est la métaphore.

2.2. L'importance du contexte

L'analyse qui a été faite ci-dessus ne tient pas compte du contexte dans lequel les fondeurs se trouvent et dans lequel le lexique est utilisé. Si nous voulons par contre étudier le vocabulaire tel qu'il est utilisé dans la vie réelle, ces aspects contextuels constituent des informations indispensables comme il a été démontré par Weinreich *et al.* (1967) et par Milroy (1980/1987). Dans la présente partie, nous nous focaliserons sur la situation sociolinguistique et les trajectoires professionnelles des fondeurs individuels, afin

d'expliquer certaines des différences que met en lumière la comparaison des vocabulaires utilisés par chaque fondeur en particulier.

Tout d'abord, certains aspects sociolinguistiques influencent directement le lexique utilisé par les fondeurs. La langue maternelle du fondeur peut par exemple influencer le vocabulaire qu'il utilise dans l'atelier. Ainsi, un forgeron touareg qui utilise le hausa dans l'exécution de la technique de la fonte de l'aluminium, emploie le mot *tamasheq* $n\alpha\lambda wa t \alpha$ au lieu des termes *pánkà* ou *zùzzúgî* – utilisés par la majorité des fondeurs hausa – pour dénommer le soufflet. Si la langue maternelle du fondeur peut influencer son propre vocabulaire, la langue utilisée par son patron peut également jouer un rôle important. La plupart des fondeurs zarma qui utilisent des emprunts au hausa, ont appris la technique chez un hausa dont ils ont hérité ces lexèmes. La langue utilisée dans l'atelier importe également : dans un atelier multilingue par exemple, les collègues fondeurs peuvent s'influencer sur le plan linguistique. Ce multilinguisme est donc également à l'origine de l'emploi des emprunts hausa en zarma. La situation sociolinguistique générale joue aussi un rôle. Hanafiou (2004 : 206-207) nous dit qu'au Niger la langue hausa est plus dominante que le zarma surtout sur le plan économique, ce qui peut expliquer qu'on trouve des emprunts au hausa dans le lexique des fondeurs zarma, tandis qu'il n'y a pas d'emprunts au zarma dans le vocabulaire des fondeurs hausa. Il faut noter aussi que les procédés de formation de mots sont plus couramment utilisés dans le lexique de la fonte en hausa qu'en zarma. Le zarma recourt plus facilement aux emprunts français pour la dénomination de ces outils. Cela peut s'expliquer par le fait que la plupart des fondeurs hausa se trouvent dans des villes comme Maradi et Zinder, où le hausa est majoritaire et qui sont géographiquement et économiquement proches des grandes villes hausaphones du Nord du Nigeria (Rouillet 2004 : 183).

La trajectoire professionnelle parcourue par le fondeur individuel peut aussi influencer son vocabulaire. Un emprunt au français très particulier a été remarqué dans le vocabulaire de certains fondeurs. Il s'agit de la dénomination du creuset, *gàmá:tî*, un terme qui provient d'un vocabulaire régional français propre aux maçons. Comment un terme de la maçonnerie est-il entré dans le lexique de la fonte de l'aluminium ? Cet emploi particulier s'explique quand on sait que l'un des utilisateurs du mot a passé plusieurs années en apprentissage chez un maçon. Les deux autres fondeurs qui utilisent le même terme sont respectivement son apprenti et son patron. Un deuxième exemple qui illustre bien l'importance des trajectoires professionnelles est l'emploi des termes de la métallurgie dans les ateliers de fonte. Chez les fondeurs hausa de Maradi, dont beaucoup font partie d'une famille de forgerons traditionnels, il est intéressant de relever plusieurs termes de la terminologie métallurgique qui ne sont pas connus par les fondeurs n'ayant pas de lien avec une famille de forgerons. L'influence du stade d'apprentissage est un dernier exemple de l'importance de la trajectoire professionnelle. Un apprenti au début de son apprentissage ne peut pas encore dénommer tous les outils qui se trouvent dans l'atelier, son vocabulaire n'étant pas aussi riche que celui de son maître.

2.3. Synthèse

À première vue, le vocabulaire de la métallurgie de récupération ne semble pas être un vocabulaire très spécialisé. Beaucoup de noms d'outils sont constitués à partir de lexèmes qui existaient déjà dans la langue concernée, soit par dérivation et composition, soit par les mécanismes sémasiologiques qui permettent que le sens d'un lexème donné soit adapté à un nouveau contexte. La prépondérance de ces « lexèmes non-indépendants » – pour utiliser la terminologie de Bostoen (2003-2004 : 204) – n'entraîne par contre pas automatiquement l'absence de spécialisation. Traitons d'abord le cas des déverbatifs. Les mots sont le plus souvent dérivés des verbes génériques, connus par chaque locuteur de la langue. Bien qu'un locuteur non spécialisé sache de quel verbe un tel déverbatif est dérivé, il n'établit pas de rapport entre le sens de ce mot et la technique de la fonte de l'aluminium. La même chose vaut pour les noms des objets utilisés dans la vie quotidienne : un non-spécialiste n'est pas au courant de l'usage exact de ces objets dans une technologie particulière. Nous pouvons donc affirmer qu'il y a une différence sémantique fondamentale entre ces verbes et leurs noms dérivés, ou entre ces mots communs et leur emploi dans l'atelier, ce qui nous conduit à conclure qu'il s'agit bien d'un lexique spécialisé.

Une deuxième caractéristique du vocabulaire de la fonte de l'aluminium est sa grande diversification, qui pourrait suggérer une dénomination ou une redénomination propre à chaque fondeur particulier. Par contre, l'analyse des contextes sociolinguistiques et des trajectoires professionnelles des fondeurs individuels montre que ces aspects-là ont une grande influence sur le vocabulaire qu'ils utilisent. Le choix que le fondeur fait d'un mot particulier n'est pas dû au hasard, mais s'explique par le contexte dans lequel il vit.

Le vocabulaire de la fonte de l'aluminium ne s'est donc pas constitué *ex nihilo*. Il se fonde sur des lexèmes qui étaient déjà présents dans la langue. Les forgerons qui ont changé d'activité et qui exécutent maintenant la technique de la fonte de l'aluminium ont adapté la terminologie de la forge aux outils qui sont utilisés par la technique nouvelle. Les deux caractéristiques décrites ci-dessus pourraient être expliquées par l'émergence récente ou par le caractère « informel » de la technique. Avant de confirmer ou infirmer ces hypothèses, nous comparerons dans la partie suivante le vocabulaire de la fonte de l'aluminium à d'autres lexiques techniques.

3. COMPARAISON DU LEXIQUE DE LA FONTE DE L'ALUMINIUM À D'AUTRES LEXIQUES TECHNIQUES

Bien qu'il s'agisse d'une étude comparative portant sur un grand nombre de langues différentes, l'étude du vocabulaire relatif à la poterie en bantou de Bostoen (2003-2004, 2005) est un bon point de départ pour cette comparaison. Quoique la technique de la poterie soit une technique beaucoup plus ancienne que celle de la fonte de l'aluminium, les lexiques des deux techniques se ressemblent sur plusieurs plans. Cette étude nous montre que le vocabulaire de la chaîne opératoire de la poterie en bantou est par exemple très diversifié et que les mêmes mécanismes du changement lexical y sont employés – comme la dérivation déverbative, la construction connective, la composition et le transfert lexical par glissement de sens. La question de la spécialisation du lexique se pose également et a été abordée par Bostoen (2005 : 314) de la manière suivante : « Suite à la modification

qu'ils ont subie, qu'elle soit d'ordre sémantique ou morphosémantique, tous ces noms deviennent en quelque sorte des noms spécialisés [...] leur utilisation n'est connue que par les artisans-potiers ».

Nous avons comparé les lexiques de la fonte de l'aluminium et de la poterie dans des langues africaines, mais que se passe-t-il si la comparaison porte sur le lexique d'une technique de caractère peu industriel dans une langue européenne? Le lexique des souffleurs de verre dans la région de Charleroi en Belgique, tel qu'il était utilisé dans les verreries jusqu'à 1930, nous semble un bon exemple. Dans l'étude de Roland (1980-1981), nous identifions les mêmes mécanismes du changement lexical. Les termes utilisés ont le plus souvent leur origine dans le langage wallon commun, bien qu'il y ait sans doute des changements sémantiques importants. Les verbes qui décrivent les actes techniques sont aussi génériques que ceux qui sont utilisés pour la fonte et la poterie; les dérivés manifestent généralement une plus grande aptitude à exprimer des sens spécialisés.

Un autre lexique auquel nous voulons comparer le vocabulaire de la fonte est celui des chemins de fer en France. Wexler (1955 : 128-129) nous fournit une étude de la formation de ce vocabulaire entre 1778 et 1842 dans laquelle il conclut : « Pendant cette période descriptive, celle que nous avons appelée "pré-mécanique", on avait donc préféré la périphrase ou le terme générique qui [...] n'était pas appelé à se spécialiser. Quand on vient à l'époque de la dénomination, quels sont les moyens linguistiques utilisés dans la formation de notre terminologie? Le vocabulaire ferroviaire montre très peu de créations de mots. [...] Les problèmes les plus délicats de la formation du vocabulaire des chemins de fer sont ceux soulevés par les nombreux mots introduits de l'anglais. Il faut surtout souligner que les raisons motivants ces emprunts sont d'ordre social. [...] À une époque légèrement postérieure, lorsque le vocabulaire fondamental est plus ou moins arrêté, et que les chemins de fer français ont acquis une indépendance certaine, on fait appel à des techniques voisines pour nommer quelques objets de moindre importance. » Afin d'illustrer les mots génériques, il donne les exemples suivants : « [N]ous avons vu plusieurs mots d'origine française employés au XVIII^e siècle là où aujourd'hui on dirait rail : ce sont *guide, bande, longuerine, tringle, plaque, barreau*. » (Wexler 1955 : 46). Il est clair que les premiers mots étaient plus génériques que le mot quotidien « rail », qui est un emprunt à l'anglais. L'auteur nous explique aussi qu'à un moment donné plusieurs expressions étaient utilisées par une personne ou par plusieurs personnes simultanément pour la dénomination d'un seul objet, ce qui est le cas pour le vocabulaire de la fonte de l'aluminium.

Même le vocabulaire maritime, un des exemples standards quand on parle d'une « vraie » terminologie, n'a pas toujours été uniforme : « Ce vocabulaire s'est formé, dès le IX^e siècle, à l'époque où se sont réunis les éléments du vocabulaire maritime de l'Occident [...] Sur les rivages de la France, ce vocabulaire très particulier a constitué deux langues distinctes jusqu'au XVI^e siècle, ayant chacune leur origine : [...] le vocabulaire du Ponant [...] et le vocabulaire du Levant [...] Ces deux vocabulaires composèrent deux langues distinctes jusqu'au XVI^e siècle. Ce n'est qu'au XVII^e que l'on peut parler de terminologies propres aux galères et aux vaisseaux. [...] Le milieu du XIX^e va marquer le moment où les techniques tendent à l'uniformisation et font ainsi disparaître les formes, les types, les procédés et les traditions dont il s'agit précisément de conserver le souvenir. » (Villain-

Gandossi 1999 : 22-23). Le vocabulaire maritime d'aujourd'hui est par contre une terminologie bien uniformisée et bien spécialisée, il y a même une grande uniformité qui dépasse les frontières linguistiques (Novi 1999 : 110). Notons aussi que le vocabulaire maritime a besoin d'une grande spécialité, comme nous l'explique la citation de Maurice d'Hartoy (1944 : 17) dans Villain-Gandossi (1999 : 21) : « Le langage des gens de mer est le plus savoureux du monde. Il est énergique et précis, abondant, harmonieux, toujours imagé. Et cela se conçoit aisément puisqu'il a pris naissance au contact des réalités souveraines, où l'homme est en péril à chaque instant, où les actes sont mâles et doivent trouver dans le mot une signification instantanée, exacte, pure de toute ambiguïté, de toute incertitude. À la mer, les minutes [...] les secondes sont précieuses ; et l'emploi malchanceux de l'une d'elles peut coûter la vie du vaisseau et de son équipage... ».

4. CONCLUSIONS

Le lexique de la fonte de l'aluminium en hausa et en zarma se caractérise tout d'abord par l'absence d'une uniformité totale entre les vocabulaires utilisés par les fondeurs individuels. Les diversités qui sont présentes ont cependant leur origine dans la situation sociolinguistique et socioprofessionnelle dans laquelle le fondeur se trouve. Une étude du contexte nous l'a montré. Le vocabulaire de la fonte s'est constitué principalement à l'aide de lexèmes non indépendants qui sont dérivés des lexèmes qui existaient avant dans la langue. Les procédés de dérivation déverbativ, de composition, et de transfert lexical sont les plus attestés. Ils produisent une spécialisation partielle, mais tout de même non contestable.

La diversification du vocabulaire de la métallurgie de récupération – bien qu'il s'agisse ici plutôt d'une diversité intralinguistique plutôt qu'interlinguistique – et sa constitution particulière apparaissent aussi dans le vocabulaire de la chaîne opératoire de la poterie dans les langues bantoues. Ces caractéristiques ne peuvent être expliquées exclusivement par l'émergence récente de la technique de la fonte, puisqu'on retrouve ces mêmes caractéristiques dans les vocabulaires beaucoup plus anciens pour lesquels celui de la poterie est un bon exemple. La comparaison du lexique de la fonte avec le vocabulaire des souffleurs de verre nous a appris que la constitution d'un lexique technique se fait d'une manière comparable dans des langues relevant de plusieurs groupes et dans des cultures très diverses. De même, le vocabulaire des chemins de fer en France permet de relever beaucoup de ressemblances. Cette comparaison nous a par contre appris que le vocabulaire d'une technique peut changer en fonction du stade de développement de la technique. Au cours du XVIII^e siècle, les chemins de fer en France étaient toujours très régionalisés tandis qu'au milieu du XIX^e siècle, ils avaient acquis un caractère national et même international. Cette évolution a eu comme résultat l'uniformisation et l'internationalisation – par des emprunts à l'anglais – du vocabulaire des chemins de fer. Le vocabulaire de la métallurgie de récupération se trouve toujours à un stade peu uniformisé, le contact régulier avec d'autres fondeurs et l'évolution historique de la technique pourraient avoir pour résultat l'uniformisation et la spécialisation, mais on a vu que, pour le vocabulaire de la poterie, ce n'est pas nécessairement le cas. Il nous semble que le caractère de l'exécution de la technique a plus d'influences sur l'évolution du lexique concerné. Bien que la soufflerie se situe dans une usine, le caractère artisanal de chaque étape de la production est conservé et

les outils employés sont restés les mêmes. Ceci peut expliquer selon nos hypothèses – qui restent de toute façon à vérifier lors d’une recherche plus approfondie – que l’origine des « termes » employés se trouve directement dans le langage quotidien. Autrement dit, c’est le contexte de l’exécution de la technique qui détermine le développement de la terminologie utilisée. Le lexique d’une technique « informelle », peu industrialisée – comme le vocabulaire de la fonte de l’aluminium, de la poterie et de la soufflerie et comme aussi le vocabulaire des chemins de fer et le vocabulaire maritime au moment de l’émergence de ces techniques – est beaucoup moins uniformisé et spécialisé que le lexique d’une technique plus « formelle » et industrialisée – comme le vocabulaire des chemins de fer à une époque plus récente – ou celui d’une technique qui s’exécute dans un monde très fermé – comme le vocabulaire maritime.

BIBLIOGRAPHIE

- Bostoën (K.), 2003-2004 : *Étude comparative et historique du vocabulaire relatif à la poterie en bantou*, Bruxelles, Université libre de Bruxelles, thèse de doctorat.
- Bostoën (K.), 2005 : *Le travail de la poterie dans le domaine bantou, une étude comparative et historique de son vocabulaire*, Frankfurt am Main, Peter Lang, Schriften zur Afrikanistik, Band 9.
- Crowley (T.), 1994 : *An introduction to historical linguistics, second edition*, Oxford, Oxford University Press.
- Geeraerts (D.), 1997 : *Diachronic prototype semantics, a contribution to historical lexicology*, Oxford, Clarendon Press.
- Hanafiou (H. S.), 2004 : « Niger » dans Chaudenson (R.) et Rakotomalala (D.), édit., *Situations linguistiques de la Francophonie, état des lieux*, Agence universitaire de la francophonie, pp. 205-209.
- Milroy (L.), 1980/1987 : *Language and social networks*, Oxford, Basil Blackwell.
- Novi (C.), 1999 : « Multilingual harmonization and standardization of technical terminology at the international maritime organization », dans Newman (D. L.) et Van Campenhoudt (M.), édit., *Terminologie maritime : traduire et communiquer, actes du 1^{er} colloque international de terminologie maritime, Bruxelles, 15 et 16 mai 1998*, Bruxelles, Éditions du Hazard, pp. 110-125.
- Roland (N.), 1980-1981 : *Le Vocabulaire du souffleur de verre à vitres dans le pays de Charleroi, étude dialectologique*, mémoire présenté en vue de l’obtention du grade de licenciée en philologie romane, Université libre de Bruxelles.
- Rouiller (F.), 2004 : « Représentations linguistiques des locuteurs natifs du hausa », dans Jolivet (R.) et Rouiller (F.), édit., *Pratiques et représentations linguistiques au Niger, résultats d’une enquête nationale*, Cahiers de l’ILSL, n° 15, pp. 179-181.
- Villain-Gandossi (C.), 1999 : « De Robert Estienne à Heinrich Paasch : la place du vocabulaire maritime dans les dictionnaires plurilingues », dans Newman (D. L.) et Van Campenhoudt (M.), édit., *Terminologie maritime : traduire et communiquer, actes du 1^{er} colloque international de terminologie maritime, Bruxelles, 15 et 16 mai 1998*, Bruxelles, Éditions du Hazard, pp. 21-46.
- Weinreich (U.), Labov (W.) et Herzog (M. I.), 1968 : « Empirical foundations for a theory of language change », dans Lehmann (W. P.) et Malkiel (Y.), édit., *Directions for historical linguistics, a symposium*, Austin/London, University of Texas Press, pp. 96-195.
- Wexler (P. J.), 1955 : *La Formation du vocabulaire des chemins de fer en France (1778-1842)*, Genève, Droz.

CLÔTURE

Allocution de clôture



Philippe Thoiron
coordonnateur du réseau LTT

Mesdames, Messieurs,

Quand nous avons choisi la thématique de ces journées scientifiques nous avons souhaité rapprocher spatialement, à l'aide d'une virgule et d'une conjonction (opportunément décrite comme « de coordination » — ce qui fait de nous deux des parents) des entités qui entretiennent, ou ont entretenu, des rapports fort différents.

Il y a belle lurette qu'on ne cherche plus à isoler le mot de son contexte. Les lexicographes, qui sont le plus enclins à traiter le mot comme un solitaire précieux, ne renoncent pas à le replacer dans sa gangue avec les citations et les exemples. Mais avec le terme c'est une autre histoire. On nous a dit et expliqué qu'il n'avait besoin que de lui-même et que son domaine est son royaume. On s'est aperçu que ce terme partait en mission et s'enrichissait (comme nous tous qui sortirons d'ici beaucoup plus riches de connaissances de toute nature et de toute forme). En s'enrichissant, on se vulgarise parfois. Le terme l'a fait. Il est devenu banal. On peut même parfois le prendre pour un mot. Et la boucle est bouclée. Voici le terme inexistant sans son contexte.

Ce que je viens de décrire (en vous donnant peut-être l'impression que je suis un coordonnateur étrange — c'est-à-dire un usager fantaisiste de la coordination conceptuelle) ce n'est pas une caricature, c'est une évolution. Et comme toute évolution elle mérite réflexion. Voilà pourquoi nous vous avons convié à ces journées scientifiques, sur ce thème.

J'ai noté que, au cours de ces journées scientifiques, on a donné à « contexte » des significations, et des objectifs, très différents ; on a évoqué la taille, avec « micro-contexte » contre « macrocontexte », mais on pourrait ajouter « collocation », « unité phraséologique ». Et bien sûr « *corpus* ». Pour les objectifs on constate que la place de la désambiguïsation est centrale. Qu'il s'agisse de calculer la valeur significationnelle en langue générale ou de lever les ambiguïtés des terminologies ou des jargons d'entreprises. Qu'il s'agisse de l'élaboration d'atlas linguistiques

(atlas linguistique de Tunisie, atlas de la Wallonie par exemple). Qu'il s'agisse de la place des informations contextuelles à associer aux informations lexicales dans les dictionnaires. Qu'il s'agisse de l'étude contrastive de vocabulaires spécialisés.

Nous avons tous remarqué que, à travers ces différentes fonctions, on atteignait des concepts fondamentaux à nos disciplines. Fondamentaux et conflictuels. J'ai retrouvé les couples infernaux traditionnels, et en particulier ceux de monosémie-polysémie, onomasiologie-sémasiologie ou encore univocité-variabilité. Mais j'ai noté avec satisfaction qu'on évoluait vers des formules plus souples, moins binairement agressives. On parle de continuum. On a évidemment accordé une large place à l'informatique et à l'automatisation, mais sans renoncer à valoriser la compétence du locuteur humain.

La qualité des communications et la courtoisie des échanges sont évidemment à votre honneur. Je me réjouis de voir que les tenants de la terminologie textuelle, les usagers fidèles des *corpus* informatisés n'ont pas jugé bon (quelle que soit la vigueur de leur conviction) de jeter les bébés de parents plus traditionnels (ou traditionalistes) avec l'eau de leur bain. Il doit y avoir place, dans un réseau comme LTT, pour des points de vue, des approches et des méthodes qui ne vivent pas de l'ostracisme ou du refus mais se renforcent, se confortent et se bonifient grâce aux liens entre les chercheurs.

C'est pourquoi le site LTT doit être utilisé pleinement. Il informe les membres du réseau (je rappelle qu'on y trouve par exemple tout ce qui concerne notre présent appel à collaboration — date de remise des dossiers : 20 octobre, ou l'appel à communication et à inscription des journées scientifiques communes de Dakar en mars 2006). Mais il peut servir aussi de forum afin que les membres échangent et discutent. Vous voyez ainsi que les suites de ces journées scientifiques peuvent se retrouver, dans les actes bien sûr, mais sur la toile aussi.

Nous établirons ainsi, ensemble, une situation dans laquelle nous pourrions mieux fédérer les actions du réseau (les actions de recherche en réseau, les journées d'animation et de formation régionale, les journées scientifiques spécifiques et les journées scientifiques communes). Nous éviterons alors les périodes d'accalmie, dont j'ai parlé ici à l'occasion de l'ouverture de ces travaux, et nous préparerons mieux les futures manifestations.

S'agissant de la présente manifestation, qui s'achève aujourd'hui, je veux renouveler, au nom du comité et — je crois pouvoir le dire — au nom de tous, nos remerciements aux autorités de l'ISTI et de la HEB, et en particulier à Daniel Blampain et à toute son équipe qui nous ont accueillis et aidés avec la convivialité et l'efficacité qui font l'envie des invités. Nous devons aussi remercier l'AUF, que ce soit à Bruxelles ou à Paris, pour sa contribution à la mise en place de ces journées.

J'espère que vous aurez apprécié ces Journées et que la lecture des actes vous remettra en mémoire des échanges fructueux établis dans une ambiance conviviale. C'est à cette tâche de confection des actes que nous allons nous consacrer maintenant. Vous voyez que le réseau continuera à s'intéresser aux mots et aux termes, même s'il s'agira cette fois d'un autre contexte.

Je vous remercie d'être venus travailler ensemble et avec nous et j'espère que ces bonnes pratiques auront maintenu et renforcé la confiance que vous faites au réseau LTT et à ses activités. Soyez en tout cas assurés que le comité et son coordonnateur feront de leur mieux pour mériter cette confiance et vous aider dans vos travaux.

Table des matières

Comité scientifique	5
Comité d'organisation	5
Membres du comité de réseau LTT	6
Appel à communications. Document de présentation (extraits)	7
Liste des auteurs.....	13
Séance d'ouverture	19
Allocution de Madame Michèle Gendreau-Massaloux, recteur de l'AUF	21
Allocution de Monsieur Philippe Thoiron, coordonnateur du réseau LTT de l'AUF.....	25
Allocution de Monsieur Daniel Blampain, professeur à l'ISTI, membre du comité LTT de l'AUF, directeur du centre de recherche TERMISTI	29
Arnaud Pierre J.L.	
<i>Phénomènes de sous-spécification sémantique : représentation des unités lexicales et prise en compte du contexte</i>	33
Anastassiadis-Syméonidis Anna et Tsaknaki Olympia	
<i>Inférences et le rôle du micro-contexte et du macro-contexte</i>	45
Alexandru Maria-Cristina et Gaudin François	
<i>Les contextes : à la source du terme ?</i>	59
Diakité Mamadou	
<i>Le trou lexical entre langue et discours : pour un statut discursif du silence</i>	69
Costa Rute	
<i>Texte, terme et contexte</i>	79
Lerat Pierre	
<i>Terme et micro-contexte. Les prédications spécialisées</i>	89
Escartin Arilla Ana	
<i>Le travail terminologique pour la traduction de l'histoire</i>	99
Elimam Abdeljlil (Abdou)	
<i>Entre prototypisation et mise en discours : les enjeux du sens</i>	109
Durieux Christine	
<i>Le contexte : filtre ou membrane ?</i>	121
Merten Pascaline	
<i>XML/XSL(T) pour un traitement unifié des corpus multilingues et des règles de traduction correspondantes</i>	129

Fafiotte Georges

Interprétariat à distance, collecte et diffusion collaboratives de corpus oraux bilingues, en situation ouverte, finalisée ou contextualisée..... 145

Mustafa El Hadi Widad

Indexation humaine et indexation automatisée : la place du terme et de son environnement..... 157

Mokhtari Abdelkrim

Marqueur, un logiciel de marquage semi-automatique de textes 167

Zimina-Poirot Maria

Topographie bitextuelle et approches quantitatives de l'extraction de ressources traductionnelles à partir de corpus parallèles..... 175

Van Campenhoudt Marc, Thierno Cissé et Muraille Paul

Élaborer des corpus XML en langues partenaires : quelles technologies appropriées ? Le cas de la Mauritanie et du Sénégal 187

Phan Huy Khánh, Vo Trung Hung et Boitet Christian

Une composition de solutions génériques pour créer le dictionnaire FEV et importer le vietnamien dans Papillon..... 199

Ben Hariz Sondos

Les contextes de l'enquête : le cas de l'Atlas Linguistique de Tunisie..... 213

Ayewa Noël Kouassi

Mots et contextes en FPI et en Nouchi..... 221

Desmet Isabel

Variabilité et variation en terminologie et langues spécialisées : discours, textes et contextes 235

Catala Dolors et Baptista Jorge

Les adverbies composés dans le domaine du travail..... 249

Blaise Jean-Yves et Dudek Iwona

Des termes au concept : de cas du vocabulaire architectural 265

Leroyer Patrick et Simonsen Henrik K.

Contexte d'utilisation, contexte de communication : la double identité du dictionnaire d'entreprise 283

Jacques Marie-Paule

L'emploi de termes réduits comme révélateur de la centralité dans le domaine 299

Fontanet Mathilde

La traduction technique : le texte sous l'empire de l'extratextuel 309

Bertaccini Franco et Matteuci Alessandra

La terminologie d'entreprise et ses contextes d'usage 317

Bellynck Valérie, Boitet Christian et Kenwright John

Construction collaborative d'un lexique français-anglais technique dans IToldU : contribuer pour apprendre 327

Ndiaye Modou	
<i>Le champ sémantique du mot marabout en français du Sénégal</i>	337
Jacquemin Bernard	
<i>Interroger un corpus par le sens : une approche linguistique</i>	347
Goffin Roger	
<i>Quels corpus et quels approches pour une description contrastive de l'eurolecte ?</i>	361
Bertels Ann	
<i>Les spécificités en contexte : comment étudier la polysémie dans un corpus technique ?</i>	371
Neveu Franck	
<i>Un aspect de l'apport des corpus à la terminologie linguistique : l'alignement</i>	381
Mathieu-Colas Michel	
<i>Les noms de divinités : web, contextes et classes d'objets</i>	391
Massoumou Omer	
<i>Les particularités lexicales dans les romans, entre discours et fiction</i>	409
Silva Raquel	
<i>Morphologie de spécialité : regard(s) sur le(s) contexte(s)</i>	421
Temmerman Rita, Kerremans Koen et Vandervoort Veerle	
<i>La termontographie en contexte(s)</i>	429
Ouerhani Béchir	
<i>La traduction de la métalangue : la problématique terme/mot en contexte</i>	441
Contente Madalena	
<i>Termes et textes : la construction du sens dans la terminologie médicale</i>	453
Barna Corina Georgeta	
<i>Approche contrastive du corpus bilingue français - roumain du domaine vétérinaire</i>	467
Balliu Christian)	
<i>Le langage de la médecine : les mots pour le dire</i>	475
Trurnit Verbic Renate	
<i>L'incidence du contexte sur la traduction médicale</i>	483
Tosheva Aneta	
<i>L'expression de la subjectivité à travers les verbes de perception visuelle dans des textes français relevant des examens cliniques et paracliniques</i>	497
Lino Teresa	
<i>Contextes et néologie terminologique dans le domaine médical</i>	509
Ilunga Ntumba Willy	
<i>La créativité lexicale du français en République Démocratique du Congo : un cas d'analyse lexicologique en contexte</i>	515
Sader Feghali Lina	

<i>La presse vue à travers Néoscope : quand les contextes médiatiques sont mis au service de la néologie</i>	525
Ravaoarimalala Élisabeth	
<i>Moderniser sa langue pour mieux apprendre les langues étrangères</i>	535
Mejri Salah	
<i>La reconnaissance automatique des néologismes de sens</i>	545
Dussart André	
<i>Critiquez vos dictionnaires bilingues... à bon escient</i>	559
Desurmont Jean Nicolas	
<i>Du contexte à la citation : les récents développements de la dictionnairique</i>	567
Dubreil Estelle et Daille Béatrice	
<i>Analyse sémantico-discursive des collocations lexicales en corpus spécialisé : la base connaissance-s</i>	579
Vaxelaire Jean-Louis	
<i>Le nom propre en contexte — une approche lexicologique</i>	591
Nzessé Ladislas	
<i>Mots et contexte : essai d'analyse lexico-sémantique du vocabulaire politique au Cameroun de 1990 à 1992</i>	599
González Rodríguez Antonio	
<i>Le dictionnaire des collocations en ligne</i>	609
Selosse Philippe	
<i>Traduire les termes de couleur : la chromonymie en botanique à la Renaissance</i>	619
Sanogo Mamadou Lamine	
<i>La recherche terminologique dans un dialecte ouvert : le cas du dioula</i>	631
Sliwa Dorota	
<i>Le terme "bien commun" et la construction du sens. Mais dans quel contexte ?</i>	641
Said Mosbah	
<i>Le stéréotype, du mot au concept : saisies à travers les contextes</i>	651
Torterat Frédéric	
<i>L'exemplification bilingue des Mots de la grammaire en contexte créolophone</i>	661
Vanghuchten Lieve	
<i>L'unité phraséologique en didactique des langues étrangères de spécialité : le cas du discours économique d'entreprise en espagnol</i>	673
Messaoudi Leila	
<i>Les rôles de la situation et du contexte dans les technolèctes bilingues français - arabe</i>	687
Mejri Soumaya	
<i>La terminologie des sciences de gestion en contexte : terme ou mot ?</i>	699
Vergely Pascale	

<i>Prépondérance du contexte extralinguistique dans la construction du sens : l'exemple des communications de travail dans la navigation aérienne</i>	709
Van der Veken Anneleen	
<i>Le rôle du contexte dans les lexiques techniques : le cas de la métallurgie de récupération en Afrique de l'Ouest</i>	719
Clôture	731
Allocution de Monsieur Philippe Thoiron, coordonnateur du réseau LTT de l'AUF	733
Table des matières	737

Une des problématiques devenues dominantes aujourd'hui en lexicologie, en terminologie et en traduction est celle des contextes. La linguistique de corpus remet en cause les frontières entre disciplines. Tout se joue dans l'analyse de corpus textuels et non plus simplement dans la pragmatique d'une phraséologie. Le contexte cesse d'avoir valeur d'illustration et représente l'élément fondamental de la construction du sens, notamment dans le traitement automatique des langues.

Le traducteur demande des corpus de référence fiables en matière de langues de spécialité. Il est à la recherche des liens pertinents entre le texte de corpus et le texte à traduire. La circulation des textes au-delà des domaines doit permettre de recentrer son activité sur les contraintes communicationnelles.

Le terminologue est amené à remettre en cause le système conceptuel propre à un domaine, le pouvoir de structuration du domaine et le logicisme ambiant au profit de pratiques textuelles réelles. Il se trouve confronté par les macrocontextes à la variabilité des terminologies, à une autre typologie des textes et à une autre acception de la normalisation.

Le lexicologue et le lexicographe doivent analyser les implications de cette linguistique textuelle, redéfinir la notion de proximité dans l'utilisation de leurs outils informatiques et poser, selon de nouveaux principes, le mot, l'exemple, la citation.

La priorité accordée au contexte donne naissance à de nouvelles procédures et à de nouvelles considérations théoriques, au carrefour des différentes disciplines. Lesquelles ? Les lexicologues, traducteurs, terminologues et enseignants réunis aux Journées scientifiques LTT de Bruxelles tentent d'y répondre ici.

La diffusion de l'information scientifique et technique est un facteur essentiel du développement. Aussi, dès 1988, l'Agence universitaire de la Francophonie (AUF), mandatée par les Sommets francophones pour produire et diffuser livres, revues et cédéroms scientifiques, a créé une collection d'ouvrages scientifiques en langue française. Lieu d'expression de la communauté scientifique de langue française, elle vise à instaurer une collaboration entre enseignants et chercheurs francophones en publiant des ouvrages, coédités avec des éditeurs francophones, et largement diffusés dans les pays du Sud grâce à une politique tarifaire adaptée.

La collection de l'Agence universitaire de la Francophonie, en proposant une approche plurielle et singulière de la science, adaptée aux réalités multiples de la Francophonie, contribue à promouvoir la recherche dans l'espace francophone et le plurilinguisme dans la recherche internationale.

Prix Nord : 55 euros

Prix préférentiel AUF - pays en développement : 20 euros

ISBN : 2-914610-31-9



eo
ac
éditions des
archives
contemporaines

