

Sous la direction de
André Clas
Université de Montréal

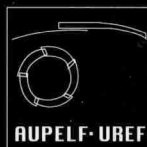
Salah Mejri
Université de Tunis

Taïeb Baccouche
Université de Tunis I

**LA MÉMOIRE
DES MOTS**
ذاكرة الكلم

Actes du colloque de TUNIS
25-26 et 27 septembre 1997

تونس 25 - 26 - 27 سبتمبر 1997



AS

actualité scientifique

SERVICED
TUNIS - TUNISIE



LA MÉMOIRE DES MOTS

ذاكرة الكلم

© AUPELF-UREF

Il est interdit de reproduire intégralement ou partiellement le présent ouvrage - loi du 11 mars 1975 - sans autorisation de l'AUPELF-UREF

ISBN : 9973 - 9753 - 5 - 9

LA MÉMOIRE DES MOTS

Vèmes Journées Scientifiques du Réseau Thématique
"Lexicologie, Terminologie, Traduction"
en collaboration avec l'Association Tunisienne de Linguistique
Tunis, Tunisie, 25 - 26 - 27 septembre 1997

Sous la direction de :

André CLAS, Université de Montréal, CANADA
Salah MEJRI, Université de Tunis I, TUNISIE
Taïeb BACCOUCHE, Université de Tunis I, TUNISIE

S E R V I C E D

TUNIS - 1998

AUPELF

B.P. 400, Succ. Côte des Neiges
Montréal (Québec)
CANADA H3S2S5

Avant-propos

La diffusion de l'information scientifique et technique est un facteur essentiel du développement. Aussi dès 1988, l'Agence Francophone pour l'Enseignement Supérieur et la Recherche (AUPELF-UREF), mandatée par les Sommets francophones pour produire et diffuser revues et livres scientifiques, a créé la collection Universités francophones.

Lieu d'expression de la communauté scientifique de langue française, Universités francophones 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 préférentielle.

Quatre séries composent la collection :

- Les manuels : cette série didactique est le cœur de la collaboration. Elle s'adresse à un public de deuxième et troisième cycles universitaires et vise à constituer une bibliothèque de référence couvrant les principales disciplines enseignées à l'université.
- Actualité scientifique : dans cette série, à laquelle appartient le présent ouvrage, sont publiés les actes des journées scientifiques organisées par les réseaux thématiques de recherche de l'UREF. A ce jour, 16 réseaux thématiques, rassemblant plusieurs milliers de chercheurs de toute la francophonie, sont constitués au sein de l'UREF.
- Prospectives francophones : s'inscrivent dans cette série des ouvrages de réflexion donnant l'éclairage de la francophonie sur les grandes questions contemporaines.
- Savoir plus Université : cette nouvelle série se compose d'ouvrages de synthèse qui font un point précis sur des sujets scientifiques d'actualité.

Notre collection, en proposant une approche plurielle et singulière de la science, adaptée aux réalités multiples de la francophonie, contribue efficacement à promouvoir la recherche dans l'espace francophone et le plurilinguisme dans la recherche internationale.

Professeur Michel Guillou
Directeur général de l'AUPELF
Recteur de l'UREF

Sommaire

Liste des auteurs	X
Membres du Comité de réseau "LTT"	XII
Discours du Pr. Dali Jazi, Ministre de l'Enseignement supérieur, à l'ouverture du colloque de linguistique "La Mémoire des Mots"	XIII
Allocution du Pr. Abderraouf Mahbouli, Président de l'Université Tunis I	XV
Présentation	1
La mémoire des séquences figées : une troisième articulation ou la réhabilitation du culturel dans le linguistique <i>Salah Mejri (ISEFC, Le Bardo, Tunisie)</i>	3
Opposés de langue, opposés de discours : les lapsus antonymiques <i>Pierre J.L. Arnaud (Université Lumière-Lyon 2, France)</i>	13
Les proverbes : des dénominations d'un type "très très spécial" <i>Georges Kleiber (Université de Strasbourg 2 et Scolia, France)</i>	23
À la recherche de la mémoire perdue. Ou pour un dictionnaire historique de l'arabe <i>Taïeb Baccouche (Université de Tunis I, Tunisie)</i>	41
Dénotation et problèmes de polysémie dans l'élaboration d'un dictionnaire électronique français-arabe <i>Bassam Baraké (Université Libanaise, Tripoli, Liban)</i>	51
Un modèle hybride pour l'extraction des connaissances : le numérique et le linguistique <i>Ismâïl Biskri, Jean-Guy Meunier, Christophe Jouis (Université du Québec à Montréal, Canada; Université Charles de Gaulle-Lille 3; Université de la Sorbonne-Paris 4, France)</i>	57
Développement de lexiques à grande échelle <i>Pierrette Bouillon, Sabine Lehmann, Sandra Manzi, Dominique Petitpierre (Université de Genève, Suisse)</i>	71
Comment représenter l'expérience individuelle qui donne leur sens aux mots, approche informatique <i>Françoise Forest (LIMSI, Orsay, France)</i>	81
Vers l'utilisation des mémoires de traduction pour la localisation des logiciels : le projet ALAMET <i>Boubaker Meddeb Hamrouni, Laurent Fischer et Mathieu Lafourcade (GETA-CLIPS- IMAG Campus; WinSoft SA, Grenoble, France)</i>	97
Quand les mots perdent la mémoire <i>Muriel Coret (Université Paris 7, France)</i>	113
L'image et la forme : aplatissement ou distorsion du temps ? <i>Xavier Lelubre (Université Lumière-Lyon 2, France)</i>	123
De l'emploi libre à l'emploi support <i>Hassane Filali Sadki (Université de Franche-Comté, Besançon, France)</i>	135
Le projet NADIA-DEC : vers un dictionnaire explicatif et combinatoire informatisé ? <i>Gilles Sérasset (GETA-CLIPS-IMAG Campus, Grenoble, France)</i>	149
TAO et théories linguistiques : institutions grammaticales <i>Philippe Blache, Jean-Yves Morin (LPL-CNRS, Nice, France; Université de Montréal, Canada)</i>	161

La mémoire des participes présent et passé <i>Poul Søren Kjærsgaard (Université d'Odense, Danemark)</i>	177
Exploration de classifieurs connexionnistes pour l'analyse de textes assistée par ordinateur <i>Jean-Guy Meunier, Ismaïl Biskri, Georges Nault, Moses Nyongwa (Université du Québec à Montréal, Canada)</i>	191
Vers la spécification et l'extension des relations terminologiques : typologie et insertion dans un système de significations des relations <i>Christophe Jouis, Widad Mustafa Elhadi, Ismaïl Biskri, Florence Le Priol (Université Charles de Gaulle-Lille 3; Université de la Sorbonne-Paris 4, France; Université du Québec à Montréal, Canada)</i>	203
Le signifié et le concept dans la dénomination <i>Marcel Diki-Kidiri (LLACAN-CNRS, Meudon, France)</i>	217
Mémoire des mots et discours : la mémoire énonciative <i>Khadiyatoullah Fall, Daniel Simeoni (Université du Québec à Chicoutimi; Université York, Toronto, Canada)</i>	223
La mémoire cinétique des termes supports <i>Amr Helmy Ibrahim (Université de Franche-Comté, Besançon, France)</i>	235
Métaphores et diachronie dans les sciences : le cas de <i>code</i>, <i>patrimoine</i>, <i>sélection</i> <i>François Gaudin (Université de Rouen, France)</i>	243
Maille ou maillon : quand des terminographes négligent l'évolution de l'usage <i>Marc Van Campenhoudt (TERMISTI, Institut Supérieur de traducteurs et interprètes, Bruxelles, Belgique)</i>	251
Les dénominations en langue de spécialité : monoréférentialité et pseudo-synonymie <i>Christine Durieux (Université de Caen, France)</i>	273
La grammaire catégorielle combinatoire applicative appliquée au français <i>Ismaïl Biskri, Jean Pierre Desclés, Christophe Jouis (Université du Québec à Montréal, Canada; Université de la Sorbonne-Paris IV; Université Charles de Gaulle-Lille 3, France)</i>	289
La forme des mots nous renseigne-t-elle sur leur nature ? <i>Sylviane Cardey, Zahra El Harouchy, Peter Greenfield (Université de Franche-Comté, Besançon, France)</i>	305
Un cas de métaphore : créativité linguistique et rôle en innovation biotechnologique <i>Myriam Bouveret (Université Paul Valéry-Montpellier 3, France)</i>	315
Les nomenclatures ont-elles une mémoire... métaphorique ? Les substrats de la nomination <i>Christine Portelance (Université du Québec à Rimouski, Canada)</i>	327
Classes d'objets et base de connaissances <i>Man-Ghyu Pak (Université de Kwandong, Corée du Sud)</i>	339
Analyse des nominalisations dans un corpus spécialisé : comparaison avec le fonctionnement en corpus "général" <i>Anne Condamines (UMR 5610, CNRS, Toulouse, France)</i>	351
La mémoire des termes : analyse conceptuelle <i>Manuel Célio Conceição (Université de l'Algarve, Faro, Portugal)</i>	369
Des mots qui ont la mémoire longue : le genou, la souris et l'arbre <i>Étienne Tiffou (Université de Montréal, Canada)</i>	377
La mémoire du commerce: aspects étymologiques et phraséologiques <i>Jeanne Dancette, Christophe Réthoré (Université de Montréal, Canada)</i>	387

Néologismes et termes d'emprunt dans le français des affaires	
<i>Jean Soubrier (INSA, Lyon, France)</i>	403
Entre stockage et réemploi en terminologie: un processus d'anamnèse?	
<i>Michèle A. Lorgnet (Université de Bologne-Forlì, Italie)</i>	419
Institutions lexicales et dictionnaires traditionnels malgaches	
<i>Jean-Yves Morin (Université de Montréal, Canada)</i>	425
Étude pour une extraction automatique de néologismes	
<i>Yvette Yannick Mathieu (Université Paris Nord-Paris 13, France)</i>	455
Repères mémoriels du lexique dans les langues bantu. Illustration à partir des structures morpho-sémantiques des substantifs kiruúndi	
<i>Jean B. Ntakirutimana (Université de Montréal, Canada)</i>	461
Les canadianismes en français et en anglais	
<i>Roda P. Roberts, Chantale Grenon-Nyenhuis (Université d'Ottawa, Canada)</i>	477
Problèmes de l'établissement de la terminologie juridique dans une langue sans officialité historique. Le cas du basque	
<i>Antton Elosegi Aldasoro (Université du Pays Basque, Donostia, Espagne)</i>	493
Organisation structurale et sémantique de la métaphore en soNay-zarma	
<i>Chaïbou Niandou (Université Abdou Moumouni, Niamey, Niger)</i>	505
'Ne m'appellez plus X mais Y' ou l'intervention du dénommé dans le processus de dénomination	
<i>René Tondji-Simen (Université de Montréal, Canada)</i>	513
Les variantes graphiques et leur source de variation	
<i>André Dugas (Université du Québec à Montréal, Canada)</i>	527
Tel métier tel mot: les conducteurs de taxi-moto "zémidjan" de Cotonou	
<i>Séverin-Marie Kinhou (Centre national de linguistique appliquée, Cotonou, Bénin)</i> ...	539
D'hier à aujourd'hui: évolution et tendances de la néologie morphologique dans le langage de l'Internet	
<i>Raymond Mopoho (Université Dalhousie, Halifax, Canada)</i>	551
Quelques nuances et une définition du mot "accent"	
<i>Charles Doutrelepoint (Université Carleton, Ottawa, Canada)</i>	563
"Motivation démotivation remotivation" ou le cycle des mots	
<i>Blandine Périchon (Université de Provence, Aix-en-Provence, France)</i>	575
Motivation et transparence des emprunts gréco-latins en français et en allemand	
<i>Christine Jacquet-Pfau, Marie-Anne Moreaux (CERTAL, Collège de France, INALCO, Paris, France)</i>	587
La dérivation en mémoire	
<i>Daniel Blampain (TERMISTI, Institut Supérieur des traducteurs et interprètes, Bruxelles, Belgique)</i>	601
Une mémoire plurielle ou une amnésie productive ?	
<i>Moufida Ghariani Baccouche (Université de Tunis I, Tunisie)</i>	607
Les mots sur le divan	
<i>Alexis Nouss (Université de Montréal, Canada)</i>	615
Les mots de l'oubli: les interjections	
<i>Emmanuel Hérique (Université de Victoria, Canada)</i>	627
La terminologie métaphorique d'Internet: aspects notionnels, structurels et traductionnels	
<i>Ingrid Meyer, Victoria Zaluski, Kristen Mackintosh, Clara Foz (Université d'Ottawa, Canada)</i>	637
Approche culturelle de la dénomination en terminologie	
<i>Atibakwa B. Edema (CELTA, Kinshasa, République Démocratique du Congo; LLACAN, Paris, France)</i>	647

Liste des auteurs

ARNAUD, Pierre J.L., Département d'Anglais Université Lumière-Lyon2, FRANCE
BACCOUCHE, Taïeb, Université de Tunis I, TUNISIE
BACCOUCHE GHARIANI, Moufida, Université de Tunis I, TUNISIE
BARAKÉ, Bassam, Université Libanaise, LIBAN
BISKRI, Ismaïl, UQAM - LANCI, CANADA
BLACHE, Philippe, LLAOR CNRS, Ministère de la recherche et de l'espace, FRANCE
BLAMPAIN, Daniel, Groupe de recherche TERMISTI, Institut Supérieur de traducteurs et interprètes, BELGIQUE
BOUILLON, Pierrette, ISSCO, Université de Genève, SUISSE
BOUVERET, Myriam, Université Paul Valéry-Montpellier FRANCE
CARDEY, Sylviane, Centre Lucien Tesnière, FRANCE
CONCEIÇÃO, Manuel Célio, Linguas e Literaturas, Unidade de Ciencias Exactas e Humanas, Universidade do Algarve, PORTUGAL
CONDAMINES, Anne, ERSS, UMR 5610 CNRS, Maison de la recherche, FRANCE
CORET, Muriel, FRANCE
DANCETTE, Jeanne, GRELT, Département de linguistique et de traduction, Université de Montréal, CANADA
DESCLÉS, Jean Pierre, CAMS-LaLIC/Université de Paris Sorbonne, FRANCE
DIKI-KIDIRI, Marcel, LLACAN-CNRS, FRANCE
DOUTRELEPONT, Charles, Faculté des arts, Département de français, Université Carleton, CANADA
DUGAS, André, UQAM, Département de linguistique, CANADA
DURIEUX, Christine, Université de Caen, UFR des langues vivantes étrangères, Esplanade de la Paix, FRANCE
EDEMA, Atibakwa B., LLACAN-CNRS, FRANCE
EL HAROUCHY, Zahra, Centre Lucien Tesnière, FRANCE
ELOSEGI ALDASORO, Antton, Faculté de Droit, Université du Pays Basque, Ramon de Lardizabal ESPAGNE
FALL, Khadiyatoulah, Université du Québec à Chicoutimi, CANADA
FILALI SADKI, Hassane, Université de Franche-Comté, FRANCE
FISCHER, Laurent, GETA-CLIPS-IMAG, FRANCE
FOREST, Françoise, Groupe Langage & Cognition, LIMSI/CNRS, FRANCE
FOZ, Clara, École de traduction et d'interprétation, Université d'Ottawa, CANADA
GAUDIN, François, Département des sciences du langage, UFR des Lettres, Université de Rouen, IRED, FRANCE
GREENFIELD, Peter, Laboratoire d'informatique de Besançon, Université de Franche-Comté, FRANCE
GRENON-NYENHUIS, Chantale, BCD, École de traduction et d'interprétation, Université d'Ottawa, CANADA
HÉRIQUE, Emmanuel, Département de langue et de littérature françaises, Université de Victoria, CANADA
IBRAHIM, Amr Helmy, Université de Franche-Comté, FRANCE
JACQUET-PFAU, Christine, CERTAL, Université Paris Nord-Paris FRANCE
JOUIS, Christophe, Université Charles de Gaulle-Lille III, B.P., FRANCE
KINHOU, Séverin-Marie, RÉPUBLIQUE DU BÉNIN
KJÆRSGAARD, Poul Søren, Institut des langues modernes, et des sciences de la communication, Université d'Odense., DANEMARK
KLEIBER, Georges, Université des sciences humaines de Strasbourg, FRANCE
LAFOURCADE, Mathieu, GETA-CLIPS-IMAG, B.P. FRANCE

LE PRIOL, Florence, CAMS (U.M. 17 CNRS), EHESS, FRANCE
LEHMANN, Sabine, ISSCO, Université de Genève, SUISSE
LELUBRE, Xavier, CRTT, Faculté des Langues, Université Lumière-Lyon2
 FRANCE
LORGNET, Michèle A., Université de Bologne-Forli, SSLIMIT, Corso Republica,
 ITALIE
MACKINTOSH, Kristen, École de traduction et d'interprétation, Université d'Ottawa,
 CANADA
MANZI, Sandra, ISSCO, Université de Genève, SUISSE
MATHIEU, Yvette Yannick, Laboratoire de linguistique informatique, Université
 Paris Nord-Paris FRANCE
MEDDEB HAMROUNI, Boubaker, WinSoft SA., FRANCE
MEJRI, Salah, Université de Tunis I, TUNISIE
MEUNIER, Jean-Guy, UQAM - LANCI, CANADA
MEYER, Ingrid, École de traduction et d'interprétation, Université d'Ottawa, CANADA
MOPOHO, Raymond, Université Dalhousie, Département de français, Halifax (N.B.),
 CANADA
MOREAUX, Marie-Anne, CERTAL, Université Paris Nord-Paris FRANCE
MORIN, Jean-Yves, Département de linguistique et de traduction, Université de
 Montréal, CANADA
MUSTAFA ELHADI, Widad, CREDO, Université Charles de Gaulle-Lille
 FRANCE
NAULT, Georges, UQAM - LANCI, CANADA
NIANDOU, Chaïbou, Département de linguistique, Université Abdou Moumouni,
 NIGER
NOUSS, Alexis, Département de linguistique et de traduction, Université de Montréal,
 CANADA
NTAKIRUTIMANA, Jean B., Faculté des lettres et sciences humaines, Département
 des langues et littératures africaines, Université du Burundi, BURUNDI
NYONGWA, Moses, UQAM - LANCI, CANADA
PAK, Man-ghyu, Département de français, Université de Kwandong, Kangnong,
 CORÉE DU SUD
PÉRICHON, Blandine, Lou Recantoun, Allée du Castellas, FRANCE
PETITPIERRE, Dominique, ISSCO, Université de Genève, SUISSE
PORTELANCE, Christine, Département de lettres, Université du Québec à
 Rimouski, CANADA
RÉTHORÉ, Christophe, GRELT, Département de linguistique et de traduction,
 Université de Montréal, CANADA
ROBERTS, Roda, BCD, École de traduction et d'interprétation, Université d'Ottawa,
 CANADA
SÉRASSET, Gilles, GETA-CLIPS-IMAG, FRANCE
SIMEONI, Daniel, York University, CANADA
SOUBRIER, Jean, INSA, FRANCE
TIFFOU, Étienne, Département de linguistique et de traduction, Université de
 Montréal, CANADA
TONDJI-SIMEN, René, GRESLET, Département de linguistique et de traduction,
 Université de Montréal, CANADA
VAN CAMPENHOUDT, Marc, Groupe de recherche TERMISTI, Institut Supérieur
 de traducteurs et interprètes, BELGIQUE
ZALUSKI, Victoria, École de traduction et d'interprétation, Université d'Ottawa,
 CANADA

Membres du Comité du Réseau “Lexicologie, terminologie, traduction”

André CLAS, Coordonnateur du Réseau LTT, Université de Montréal, Montréal, Canada

Anna ANASTASSIADIS-SYMÉONIDIS, Université Aristote de Thessalonique, Thessalonique, Grèce

Roger GOFFIN, Université Libre de Bruxelles, Bruxelles, Belgique

Antoine LIPOU, Université Marien Ngouabi, Congo-Brazzaville

Salah MEJRI, Université de Tunis I, Tunis, Tunisie

Philippe THOIRON, Université Lumière. Lyon 2 , France

**DISCOURS DU PR. DALI JAZI
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR
A L'OUVERTURE DU COLLOQUE DE LINGUISTIQUE
"LA MEMOIRE DES MOTS"**

(Tunis, le 25 Septembre 1997)

Mes chers Collègues,
Mesdames Messieurs,

Ces Journées organisées par le Réseau Lexicologie, Terminologie, Traduction de l'AUPELF-UREF avec la collaboration de l'Université de Tunis I et de l'Association Tunisienne de Linguistique portent sur le thème "La mémoire des mots" .

Permettez-moi à cette occasion de rappeler que la Tunisie est l'un des pays fondateurs de l'AUPELF. Elle a participé régulièrement à ses instances et à ses diverses activités dans le monde. Ce n'est donc pas par hasard que ces Journées soient pour l'année en cours la quatrième manifestation de l'AUPELF-UREF en Tunisie après le Conseil d'Administration de l'AUPELF-UREF, le séminaire régional du monde arabe, les assises sur l'enseignement du français et en français.

Cette coopération a abouti à plusieurs réalisations telles que les projets de recherche favorisant la mobilité des chercheurs, les bourses d'excellence et les bourses de doctorat dont bénéficient les jeunes chercheurs, la création du centre Syfed qui met à la disposition des universitaires et des étudiants chercheurs des banques de données bibliographiques et qui facilite l'acquisition des documents disponibles dans les universités étrangères. Ainsi, un réseau de relations s'est créé entre les universités tunisiennes et les autres universités faisant partie de l'AUPELF, donnant lieu à des manifestations internationales de cette envergure.

Votre colloque, faut-il le rappeler, est le premier colloque de linguistique de cette importance qui se tient, grâce à cette coopération, en Tunisie. Il est d'autant plus important qu'il porte sur la mémoire, à travers la mémoire des mots.

C'est un thème qui pose une problématique à la fois actuelle et historique, générale et particulière, linguistique et culturelle.

Elle se situe en même temps au carrefour des spécialités linguistiques du moment qu'elle implique aussi bien le lexical que le syntaxique, le morphologique que le sémantique, la langue que le discours. C'est aussi un point de jonction entre les sciences du langage et les sciences humaines.

La langue, en tant qu'outil de communication, n'est-elle pas le lieu privilégié de la mémoire à la fois individuelle et collective, le creuset des différentes cultures et le moyen de conservation du patrimoine humain ? N'est-elle pas considérée comme un fait indissociable de l'intelligence et de la pensée humaines puisque c'est en elle et avec elle que

se conçoivent tous les discours assurant la médiation entre nous et le monde, entre l'homme et ses semblables, entre la personne et elle-même? N'est-elle pas aussi le moyen idéal de l'expression de notre affectivité ?

La langue est également le siège d'une mémoire chargée d'histoire ; elle est la mémoire qui participe à la survie de l'espèce, puisque tout être sans mémoire est voué à la disparition. Les codes génétiques, le génome humain ne sont-ils pas autant de mémoires dont la raison d'être est la préservation de l'espèce ?

La tenue de ce colloque est un événement scientifique qui permet aux spécialistes des différents pays de confronter les résultats de leurs travaux, d'exposer les fruits de leurs recherches et de faire le point des acquis réalisés dans le domaine.

L'importance de la tenue de ces journées scientifiques réside aussi dans l'impact qu'elles peuvent avoir sur le développement de la recherche fondamentale et appliquée dans la discipline et sur l'évolution des méthodes employées dans des domaines aussi variés que l'enseignement des langues, la didactique des matières, la traduction, le traitement automatique du langage, etc.

C'est un moment important pour la communauté scientifique parce que de telles rencontres favorisent les échanges, encouragent la production scientifique et aident à la conception de projets communs à des laboratoires et à des institutions dans différents pays.

Que la mémoire linguistique soit de nature phonétique, morphématique, lexicale ou sémantique, elle est toujours essentielle parce qu'elle a une dimension ontologique, en ce sens qu'elle assure la continuité, qu'elle conserve l'acquis et façonne l'identité, et que sans elle l'intelligence et les capacités de traitement ne peuvent pas exister. La révolution informatique que nous vivons actuellement, avec ses diverses applications (la robotique, les autoroutes de l'information, etc.), ne peut-elle pas être ramenée aux nouvelles formes de mémoire aux capacités de plus en plus grandes ?

Mesdames, Messieurs,

Ces journées scientifiques me procurent une occasion supplémentaire pour insister sur la place de plus en plus grande que doit occuper la recherche scientifique dans notre vie universitaire. Il n'y a de progrès de l'enseignement que par la recherche. Il n'y a de développement d'un pays que par la qualité de l'enseignement et de la recherche scientifique. Monsieur le Président de la République nous y incite sans cesse et nous procure des moyens de plus en plus importants à cette fin. A nous, universitaires, incombe la mission d'agir, de multiplier les initiatives, de développer l'esprit scientifique et la culture de l'excellence.

Je voudrais, par ailleurs, insister sur la place que devraient prendre, de plus en plus, les sciences humaines dans les programmes de recherche dans notre pays. Ce domaine est des plus importants par son objet, sa richesse, sa complexité, les satisfactions intellectuelles qu'il procure, mais aussi par sa dimension humaine. Soyez assurés que le Ministère de l'Enseignement Supérieur ne manquera pas d'accorder à la recherche en Sciences Humaines un intérêt toujours plus grand.

Je souhaite à vos travaux plein succès et à vous tous un bon séjour en Tunisie.

DISCOURS DU PR. ABDERRAOUF MAHBOULI PRÉSIDENT DE L'UNIVERSITÉ DE TUNIS I À LA CLÔTURE DES JOURNÉES

(Tunis, le 27 septembre 1997)

Monsieur le Ministre de l'Enseignement Supérieur
Mes Chers Collègues,
Mesdames et Messieurs.

Je voudrais, tout d'abord, au nom de l'Université des Lettres, des Arts et des Sciences Humaines de Tunis, souhaiter, très cordialement, la bienvenue en Tunisie à tous nos collègues ici présents.

Je voudrais également vous dire que je suis particulièrement heureux que l'Université de Tunis I soit associée, en la personne de plusieurs parmi ses enseignants-chercheurs les plus éminents, aux travaux des 5èmes journées scientifiques du réseau thématique «Lexicologie - Terminologie - Traduction».

En sa qualité de membre de l'AUELF-UREF, l'Université de Tunis I participe déjà activement aux programmes et aux instances de l'AUELF, que ce soit dans le cadre du Conseil d'Administration, des réseaux institutionnels ou des réseaux thématiques. Je me réjouis, donc, de voir les chercheurs de notre université, et grâce à leur participation à cette importante manifestation, davantage impliqués dans les programmes de l'AUELF-UREF. Et je suis persuadé que le choix du réseau LTT de tenir ses 5èmes journées scientifiques à Tunis contribuera à renforcer les liens entre nos universitaires et la communauté scientifique internationale dans un domaine aussi important que celui des études linguistiques.

Je me dois à cet égard de souligner le développement remarquable que connaissent actuellement les études linguistiques en Tunisie. Cet essor est illustré par les différentes manifestations qu'organisent les associations scientifiques telles que la dynamique Association Tunisienne de Linguistique, ou bien, l'Association de Lexicologie Arabe de Tunisie; par les nombreux travaux de grande qualité que publient régulièrement les chercheurs tunisiens dans ce domaine ou encore, et dans un cadre plus académique, par la mise en place au sein de l'Université de Tunis I, et à partir de la présente année universitaire, d'un DEA et d'une formation doctorale en linguistique.

Cet intérêt pour les études linguistiques n'est pas, à mon avis, étranger à l'une des problématiques et à l'un des défis majeurs qui se posent aujourd'hui à notre société tunisienne et qui réside dans cette conciliation, nécessaire, quoique, parfois, difficile, entre identité et modernité.

Dans la mesure où la langue participe à la construction de notre identité, à la structure de notre intelligence, de notre vision du monde, de notre inconscient même, les études

linguistiques contribuent, de ce fait, à la compréhension de notre nature fondamentale et des éléments constitutifs de notre nation.

Et c'est bien à cette réflexion sur l'identité que nous invite le thème que vous avez choisi pour vos journées scientifiques, à savoir « la mémoire des mots », ces mots qui sont les lieux de mémoire, le creuset où s'enracine l'identité nationale.

On définit, aujourd'hui, la nation comme étant le lieu qui rassemble des hommes, non pas sur ce qu'ils sont, mais sur la mémoire de ce qu'ils ont été; c'est le lieu d'une commune histoire, de communs malheurs et de communs bonheurs. Et, ce sont souvent, précisément, les mots qui traduisent, mieux que tout autre, cette histoire commune, ce destin partagé.

Identité, mais, également, modernité. Et je voudrais, ici, souligner que le développement des études linguistiques, l'enseignement et la recherche en traduction, le traitement automatique du langage comme vous vous en êtes préoccupés lors de votre séminaire de formation, sont le signe d'une volonté d'ouverture à la modernité et d'une aspiration au progrès.

Il est, d'ailleurs, significatif que c'est au cours des périodes considérées comme les plus brillantes de notre histoire arabo-musulmane que l'entreprise de traduction a connu son essor le plus remarquable. Que l'on songe, seulement, à l'École de Traduction de Tolède, ou à Beit el Hikma fondée à Bagdad par le Calife Al Mamoun, fils de Haroun al Rachid, et où s'affairait une armée de traducteurs, assoiffés de savoir.

Dans la Tunisie d'aujourd'hui, l'Académie des Sciences, des Lettres et des Arts, la nouvelle «Beit el Hikma», a repris le flambeau de la traduction; dans les écoles, la loi du 29 juillet 1991 relative au système éducatif prévoit, dans son article 1er que «le système éducatif doit permettre à chaque élève de maîtriser une langue étrangère au moins de façon à lui permettre d'accéder aux productions de la pensée universelle, à la technique, aux théories scientifiques et aux valeurs humaines et le préparer à en suivre l'évolution et à y contribuer d'une manière propre à réaliser l'enrichissement de la culture nationale et son interaction avec la culture humaine universelle». A l'Université, enfin, les enseignements s'ouvrent, non seulement aux langues vivantes étrangères, mais également aux langues anciennes et c'est à l'initiative même de notre Ministre de l'Enseignement Supérieur qu'a démarré au sein de notre université, depuis l'année écoulée, une maîtrise de latin et de grec que je considère comme un nouvel aspect de cette ouverture sur la modernité en même temps qu'un facteur d'approfondissement de notre patrimoine historique.

Mesdames et Messieurs,

La Tunisie a la chance d'être un pays homogène linguistiquement et culturellement. Mais son histoire, faite de synthèses, mille fois recommencées, ainsi que son environnement immédiat, lui imposent d'être attentive aux autres cultures, autant pour s'y ressourcer elle-même, que pour apporter, comme elle l'a toujours fait durant sa longue histoire, sa contribution à l'enrichissement de la civilisation humaine. C'est pourquoi je formule le vœu que votre présence en Tunisie, sur cette terre à l'histoire trois fois millénaire, à la mémoire si riche et si dense puisse vous inspirer et vous guider dans votre réflexion sur la mémoire des mots.

PRESENTATION

La Mémoire des Mots est, de l'avis de tous les participants aux Vèmes journées scientifiques du réseau LTT de l'AUPELF -UREF organisées à Tunis les 25, 26 et 27 septembre 1998, un thème dont le choix est très judicieux.

La formulation de ce titre que nous devons au professeur André Clas, coordonnateur du réseau LTT, réalise un raccourci tel qu'il englobe pratiquement tous les domaines de l'étude du lexique : morphologie, sémantique, emplois discursifs, traitement automatique, charge culturelle, etc.

Par sa simplicité, elle aide à surmonter les divers types de cloisonnement entre les disciplines.

C'est aussi la netteté du concept qui a favorisé un balayage systématique de tout le spectre notionnel du mot :

- des mots simples "qui ont la mémoire longue : le genou, la souris et l'arbre" (E. Tiffou) «aux dénominations d'un type «très très spécial»» que sont les proverbes (G. Kleiber) en passant par "les mots de l'oubli : les interjections" (E. Hériqué);
- des séquences figées qui représentent un vrai réceptacle de la mémoire (S. Mejri) aux mots qui "perdent la mémoire" (M. Coret);
- des mots grammaticalement marqués comme "les participes présent et passé" (P.S. Kjdergaard) aux mots du discours véhiculant "une mémoire énonciative" (K. Fall et D. Simeoni) en passant par "les mots sur le divan " (A. Nouss).

Une très grande diversité de mémoires ont été revisitées : mémoire cinétique (A.H. Ibrahim), mémoire perdue (T. Baccouche), mémoire métaphorique (C. Portelance), mémoire plurielle (M. Ghariani Baccouche), mémoire graphique (A. Dugas), mémoire des termes (M.C. Conceição), mémoire du commerce (J. Dancette et C. Réthoré), etc.

A la diversité des mémoires correspond celle des domaines de recherche :

- La terminologie est l'un des domaines privilégiés par les intervenants, qui se sont intéressés entre autres à la dénomination vue sous l'angle de l'opposition signifié / concept (M. Diki Kidiri), de la relation entre dénommé et dénomination (R. Tondji-Simen), de l'évolution de l'usage (M. Van Campenhoudt), du fonctionnement des mécanismes métaphoriques (F. Gaudin / C. Niandou), de la pseudo-synonymie (C. Durieux), de la dimension culturelle (A.B. Edema), et des différents domaines d'emploi comme le domaine juridique (A.E. Aldasoro), l'internet (R. Mopoho, I. Meyer, V. Zaluski, K. Mackintosh et C. Foz), la biotechnologie (M. Bouveret), etc.
- La traduction et le contact des langues représentent un autre axe qui a fait l'objet de plusieurs travaux en rapport avec la lexicographie (B. Baraké / J-Y. Morin), l'informatique (B.M. Hamrouni, L. Fischer et M. Lafourcade), les "institutions grammaticales" (P. Blache J-Y. Morin), l'emprunt (J. Soubrier / C. Jacquet-Pfau et M.A. Moreaux), les spécificités lexicales (R. Roberts, C.Grenon-Nyenhuus).
- L'informatique, étant un domaine de prédilection, a couvert des sujets aussi différents que :
 - .les modèles pour l'extraction des connaissances (I. Biskri, J.G. Meunier, C. Jouis);
 - .les approches de représentation de l'expérience individuelle (F. Forest);
 - .l'informatisation de certains dictionnaires (G. Sérasset);
 - .l'analyse des textes assistée par ordinateur (J-G. Meunier, I. Biskri, G. Nault, M. Nyongwa);
 - .les relations terminologiques (C. Jouis, W. Mustafa Elhadi, I. Biskri, F. Je Priol);

- .les classes d'objets (M-G. Pak);
- .l'extraction automatique des néologismes (Y-Y . Mathieu);

Nous n'avons là que les axes les plus marquants. Les lapsus (P.J.L. Arnaud), la grammaire catégorielle (I. Biskri, J.P. Desclés, C. Jouis), la morphologie (S. Cardey, Z. El Harouchy, P. Greenfield), les substantifs kirundi (J.B. Ntakirutimana) les nominalisations (A. Condamines), les termes de métiers (S.M. Kinhou), le cycle des mots (B. Périchon), l'image et la forme (X. Lelubre) sont autant d'autres sujets traités à l'occasion des Vèmes journées .

Encore faut-il préciser que les textes qui figurent dans ce volume ne représentent pas la totalité des travaux : un numéro spécial (n° 35) de la Revue Tunisienne des Sciences Sociales¹ renferme une partie des actes des Journées (plus de 400 pages).

La réussite des Vèmes Journées ne réside pas seulement dans la qualité des débats scientifiques connus par les divers ateliers; elle englobe aussi les contacts entre les chercheurs et les échanges entre les représentants des divers laboratoires et institutions universitaires qui ont eu lieu à cette occasion.

Parmi les projets conçus lors de ces Journées, nous retenons la création d'une association internationale de linguistique qui s'appelle Rencontres Linguistiques Méditerranéennes et qui aura pour siège le CERES à Tunis (voir l'adresse en note). Fondée par les professeurs Taïeb Baccouche (Université de Tunis I), André Clas (Université de Montréal), Gaston Gross (Université de Paris 13) et Salah Mejri (Université de Tunis I), cette association s'inscrit dans la perspective du traitement automatique du langage ².

Nous ne pouvons finir cette présentation sans remercier :

- les membres du réseau L.T.T, en particulier le professeur André Clas, son coordonnateur, pour avoir accepté d'organiser à Tunis les Vèmes Journées en collaboration avec l'ATL;
- le professeur Gaston Gross, Directeur du Laboratoire Linguistique Informatique de Paris 13, qui a assuré avec son équipe le séminaire de formation sur le dictionnaire électronique;
- Monsieur le Ministre de l'Enseignement Supérieur qui a bien voulu inaugurer les Journées, Monsieur le Ministre du Tourisme qui a beaucoup contribué au succès de cette rencontre scientifique, Monsieur le Secrétaire d'Etat à la Recherche Scientifique et à la Technologie pour son soutien efficace et continu, Monsieur le Maire de Tunis pour son accueil chaleureux, Monsieur le Directeur du CERES qui n'a ménagé aucun effort pour nous encourager, Madame la Directrice de l'ISEFC qui n'a pas hésité à héberger les travaux du séminaire de formation, Messieurs les Professeurs membres du comité d'honneur qui ont bien voulu parrainer nos Journées et Messieurs les Professeurs membres du comité scientifique qui ont évalué les projets des communications.
- La compagnie Tunis-Air qui a attribué aux participants des tarifs spéciaux.

Un dernier mot : au nom de tous les participants, nous adressons un remerciement «très, très spécial» aux membres de l'équipe d'organisation qui ont tout fait pour que ces Journées soient aussi agréables.

Salah Mejri - Taïeb Baccouche

1) Disponible au Centre d'Etudes et de Recherches Economiques et Sociales, 23 rue d'Espagne, 1000, Tunis (Tunisie)

2) La 1ère RLM aura lieu à Tunis les 17-18 et 19 septembre 1998 et aura pour thème "Le Figement Lexical".

LA MÉMOIRE DES SÉQUENCES FIGÉES : UNE TROISIÈME ARTICULATION OU LA RÉHABILITATION DU CULTUREL DANS LE LINGUISTIQUE

Salah MEJRI

Université de Tunis I, Tunis, Tunisie

D'habitude, quand on parle de mémoire, on pense surtout à des faits de nature abstraite, n'ayant pas forcément d'appui matériel immédiat, c'est-à-dire en l'absence de tout contrat sémiotique établi directement entre un support concret et un ensemble d'évocations plus ou moins nettes. Si nous parlons de contrat sémiotique, c'est parce qu'à l'origine de cette relation, il n'existe pas de consubstantialité entre le support et son contenu : le même support peut être en rapport avec plusieurs contenus différents. Aux antipodes de ce genre de relation se trouve celle qui fonde le signe linguistique simple, lequel fonctionne selon un mode relativement arbitraire, du moins synchroniquement. Tel ne semble pas être le cas des unités lexicales construites, particulièrement les séquences polylexicales, c'est-à-dire les séquences figées.

1. LES PARTICULARITÉS DES SF

Elles sont situées à mi-chemin entre le fonctionnement symbolique fondé sur le principe de la prégnance forte et celui des unités fondamentalement arbitraires. Comparées à ce dernier type de signes, elles se distinguent par des caractéristiques saillantes qui en conditionnent le fonctionnement :

- elles ont un signifiant lourd qui crée une dissymétrie par rapport au signifié et qui modifie en même temps les rôles respectifs des facettes de ce signe linguistique. Comparez, par exemple, *mourir* et *boire le bouillon d'onze heures* ou *casser sa pipe*; *s'enfuir* et *prendre la poudre d'escampette*; etc.;

- elles se trouvent investies d'un fonctionnement qui rappelle en même temps son origine discursive (une sorte de fonction évidentielle) et qui en fait une unité lexicale ayant un sens global propre et un signifié globalisé.

Or ce signifiant globalisé demeure remotivable, quelle que soit la soudure des constituants de la séquence. Pourquoi ?

- parce que sa nature polylexicale ne sert pas de signifiant étanche au sens, c'est-à-dire segmentable seulement en unités de la deuxième articulation (les phonèmes) ou des unités de prononciation (des syllabes);

- parce que sa polylexicalité a pour corollaire systématique une nouvelle fonction : puisque la SF emprunte son signifiant à d'autres unités du lexique et qu'elle prend naissance dans une réalisation discursive, elle porte nécessairement en elle sa propre mémoire :

- une mémoire syntaxique qui se vérifie surtout au niveau des archaïsmes : *Adviene que pourra; Plus fait douceur que violence; etc.*;
- une mémoire lexicale qui conserve plusieurs unités qui ne doivent leur existence actuelle qu'aux expressions qui les véhiculent : *au fur et à mesure*;
- une mémoire sémantique qui participe à la structuration sémantique du lexique telle qu'elle apparaît dans les articles de certains dictionnaires, par exemple.

En d'autres termes, la SF est susceptible de deux décodages : l'un, le plus courant, est fondé sur son fonctionnement en tant que signe global, et l'autre, toujours possible, qui substitue à cette approche synthétique une autre analytique remontant les constituants aussi bien au niveau de leurs signifiés que de leurs signifiants pris isolément :

- «Ils s'en allaient, têtus, sur les six bottes, continuer une enquête qui *ne tournait pas rond* comme leurs képis» (H. Bazin, *L'huile sur le feu*, Grasset 1954 : 146).
- «Plus tard, il [=Picasso] fit son portrait. Jacob lui lut ses vers et Picasso — qui *parlait français comme un peintre espagnol* — lui déclara qu'il était le seul poète français de l'époque» (*Le Monde* du 13-8-1994 : 11).

2. LES ARTICULATIONS DANS LES SF

Ces emplois sont loin d'être exceptionnels (cf. le langage journalistique, certains textes de romanciers ou poètes modernes et contemporains, etc.) et illustrent la coexistence de deux types de découpage :

- l'un est fondé sur la base du signe global (signifiant et signifié) : c'est le découpage conforme aux SF considérées comme des unités lexicales ayant un signifiant vu dans sa globalité et un signifié global correspondant :

«Il est hors de conteste, répondit M. de Norpois, que la déposition du colonel devenait nécessaire pour peu que le gouvernement pensât qu'il pouvait bien y *avoir là anguille sous roche*» (M. Proust, *À la recherche du temps perdu*, II : 240).

- l'autre est fondé sur l'autonomie des constituants avec ou sans respect de l'intégrité du signifiant :

- dans l'exemple de *ne pas tourner en rond*, de H. Bazin, la SF, bien qu'employée intégralement, sa globalité sémantique n'est pas respectée;

- dans le suivant, il est porté atteinte à la solidarité des constituants :

«C'est bien ce que je pensais, tu *caches sous roche* une mauvaise excuse» (R. Queneau, *Le vol d'Icare*, Gallimard 1968 : 94).

Donc unité et pluralité donnent lieu à des segments s'articulant soit à un premier niveau, c'est-à-dire celui du sens global, soit à un second, c'est-à-dire celui qui correspond aux unités lexicales servant de constituants à la SF; ce qui signifie qu'au niveau de la première articulation, on a affaire à une articulation impliquant des unités douées de sens à deux niveaux : une première articulation globale et une première articulation plurielle.

Cette articulation à deux étages provient d'une segmentation objective imposée au décodeur dans les situations de remotivation; ce qui en fait une donnée empiriquement vérifiable.

Remarque : on pourrait postuler un dédoublement au niveau de la deuxième articulation mais cela ne représente, bien qu'il existe dans le cas de la désintégration du signifiant (cf. les exemples de *parler français comme une vache espagnole* et *anguille sous roche*), ni un fait systématique ni une opération impliquant uniquement le signifiant de la séquence (cf. les jeux de mots sur les paronymes). C'est un phénomène qui demeure, somme toute, accidentel et très limité.

André Martinet précise dans *La linguistique synchronique* (1970 : 33) que l'importance de la double articulation ne réside pas seulement dans son caractère très économique, mais aussi dans le fait que la langue, «en confiant le soin de former ses signifiants à des unités sans face signifiée, les phonèmes, elle les protège contre les atteintes du sens». C'est pour cela que tout parallèle dans le fonctionnement entre les deux articulations ne nous semble pas fondé. Dans un cas, le sens est la raison d'être des unités; dans l'autre, l'absence de sens est la condition *sine qua non* du fonctionnement des phonèmes en tant qu'éléments constitutifs des signifiants.

Devant l'importance des manipulations auxquelles le signifiant de la SF pourrait donner lieu, on ne peut pas s'empêcher de relever la systémicité de la double segmentation opérée au niveau de la première articulation :

1- Cette systémicité provient du fait que toute séquence discursive est susceptible de faire l'objet d'une fixité totale ou partielle; les séquences figées ne sont pas nécessairement des syntagmes bien formés. Il arrive dans certains cas que les éléments fixes soient :

- des éléments de nature fonctionnelle : à la + nom ou adjectif ethnique; *tout comme*, etc. «un délirant univers à la Aldous Huxley où tout est travail, ordre et propreté» (*Le monde diplomatique*, Juin 1992 : 1).
- des phrases entières : *Les carottes sont cuites; le coeur n'y est pas*; etc.
- des formules du genre : à vos souhaits; *marché conclu*; etc.

Certains écrivains n'hésitent pas à agglutiner à l'écrit certaines séquences libres pour en faire une unité, imitant en cela les mécanismes du figement :

«À l'étranger, il faut toujours montrer un visage serein, *un pays-où-tout-se-passe-bien*. Le linge sale ne se livre pas aux yeux des étrangers» (T.B. Jelloun, *La nuit de l'erreur*, Seuil 1997 : 102).

2- Cette systémicité se traduit aussi par l'impact que le figement a sur toutes les dimensions du système :

- lexicalement, il nourrit régulièrement le lexique d'unités toujours plus expressives (cf. les argots), de séquences de plus en plus nouvelles (cf. le mouvement néologique), et d'unités phraséologiques fixant les dimensions culturelles de la communauté dans la langue (cf. par exemple les parémies);

- syntaxiquement, il dote la langue de nouveaux outils exprimant les diverses relations syntaxiques (cf. les morphèmes discontinus, les locutions prépositionnelles et conjonctives, l'expression de l'intensité, les périphrases verbales, les verbes supports, etc.);

- morphologiquement, il détermine la formation polylexicale d'un très grand nombre d'unités (cf. les locutions verbales, les structures des séquences nominales, etc.);

- sémantiquement, il participe à la structuration de la polysémie des unités lexicales (cf. par exemple, la manière dont sont construits les articles de dictionnaires comme le *G.L.L.F.* où l'organisation des sémèmes se fait sur la base des emplois figés), et fixe des contenus culturels dans le sémantisme lexical.

3- Cette systémicité est fondamentale pour :

- la survie du système puisqu'elle permet, en plus des points évoqués plus haut, d'établir une sorte de pont entre la langue et la parole (ou le discours) faisant de cette dernière un outil de renouvellement et de réaménagement successifs et constants du système : par le biais du discours, le système se réorganise et se réadapte continuellement;

- la redynamisation du système : en plus des dimensions syntaxiques et sémantiques déjà évoquées, nous rappelons l'activité terminologique qui met en relief la fonction référentielle du langage et où les SF jouent un rôle déterminant (cf. ce que dit Benveniste de la synapsie);

- l'économie générale du système : en plus des avantages certains de la double articulation du langage, elle ajoute un niveau intermédiaire de segmentation faisant d'unités initialement autonomes des formants ayant nécessairement le statut d'éléments phoniques étanches au sens, du moins dans l'interprétation globale de la séquence. C'est dans ce cadre que nous parlons du figement en tant que recyclage de la parole usée. Le système, ne se contentant plus des ressources des deux articulations, produit des SF dès qu'il est mis en usage et qu'il acquiert le dynamisme vital aux langues utilisées. C'est ce qui fait du figement un phénomène propre aux langues naturelles. Même si l'orthographe garde les traces de l'autonomie initiale, dans la conscience linguistique des locuteurs le signifiant lexical est appréhendé d'une manière globale. En d'autres termes, l'analyse des SF, appliquée à une séquence comme *noyer le poisson*, donne lieu à un découpage comme suit :

a / [nwajelɔpwasɔ]

«entretenir la confusion pour duper ou lasser son adversaire»

S_a global

S_é global

où il n'y a pas la moindre correspondance entre les constituants et la synthèse sémantique. Ainsi les unités douées de sens abandonnent-elles leur statut morphématique au profit du fonctionnement phonématique de leurs propres constituants; cet abandon n'étant jamais définitif, la réactivation de la segmentation morphématique reste toujours possible comme c'est le cas dans l'exemple suivant :

b / «Et toi, le coupa brusquement l'inspecteur Ali, tu essaies de noyer le poisson alors que l'eau est son élément.» (D. Chraïbi, *L'inspecteur Ali à Trinity college*, 1996 : 90)

où le premier découpage reste valable dans une lecture qui se limiterait à la première proposition. Dès qu'on entame la proposition adversative avec ce qu'elle contient comme reprise sémique (*noyer, poisson / eau*) et renvoi anaphorique au constituant nominal de la séquence verbale (*noyer le poisson / son élément*), s'opère une nouvelle construction sémantique qui s'ajoute à la synthèse sémantique et où les constituants de départ reprennent leur autonomie grâce à la segmentation supplémentaire évoquée plus haut. L'aspect contradictoire de cette expression dégagé à la suite de ce commentaire autonymique, ne serait pas possible sans le repli mémoriel inscrit dans la structure même de la séquence. Pour schématiser, on pourrait dire que les SF, quels qu'en soient la structure et l'emploi dans le discours, sont susceptibles d'une analyse traduisant à la fois le dédoublement et le parallélisme entre les types de S_a et les deux types de $S_é$ correspondants :

signifiant global	et	signifiant pluriel (polylexical)
signifié de synthèse		signifié analytique

Nous disons bien dédoublement non polysémie parce que nous croyons que le traitement des SF en termes de polysémie ne tiendrait pas compte de sa dimension polylexicale et l'assimilerait à une unité unilexicale. La polysémie, croyons-nous savoir, exige un cadre formel restreint (base, affixe, catégorie syntaxique...) et une filiation dans ce cadre, même s'il y a variation du comportement syntaxique en rapport avec le changement sémantique. Michel Bréal a bien perçu cette particularité dans son *Essai de sémantique* (1897) quand il a établi un lien direct entre la polysémie et la notion de mot. L'invariance du signifiant limité au cadre du mot est ce qui favorise la superposition des sémèmes entre lesquels on peut dégager des relations sémantiques diverses (cf., par ex., le traitement de la polysémie dans R. Martin, *Pour une logique du sens*, 1992).

Appliquer l'approche polysémique à la SF serait très réducteur et fausserait dès le départ les données de l'analyse :

a. Le premier problème serait celui du rattachement de la polysémie à un support signifiant précis. Puisque la SF est par définition polylexicale, rattacher ce sens à l'un de ses constituants nous conduirait à faire abstraction des autres éléments alors qu'ils participent à la synthèse sémantique de la séquence de la même manière (cf. les analyses des locutions verbales de Hervé Curat 1982 et 1984 et de A.M. Schmid 1989 et 1991).

b. Le deuxième problème serait celui de la nature de la segmentation du signifiant de la séquence : l'invariance du signifiant du mot simple repose clairement sur une segmentation conforme aux deux articulations courantes. Traiter la SF de la même manière conduirait à faire abstraction de l'articulation supplémentaire décrite plus haut.

c. Le troisième problème serait celui du rapport existant entre les unités de départ et la nouvelle signification qui est fondé dans la SF, contrairement à ce qui se passe dans les unités simples où il y a consubstantialité entre le S_a et les $S_é$ anciens ou nouveaux, sur une sorte de dédoublement : les unités de départ prêtent leurs signifiants à la séquence sans

que cette opération les prive de recouvrer leur fonctionnement libre. Elles ont donc, dans le cadre de la SF, une double attache, celle qui les relie au signifiant global dans lequel elles perdent momentanément leur identité morphématique (deuxième articulation) et celle qui assure la réactivation de l'identité temporairement perdue (l'articulation supplémentaire de la première articulation). Ne pas tenir compte de la particularité de ces liens, c'est trahir le fonctionnement réel de ce type de signes linguistiques qui participent à la fois du mot (ce qui fait leur unité) et du syntagme ou de la phrase (ce qui leur donne leur identité). Rappelons que si on parle de consubstantialité dans le cas de l'unité simple, c'est parce que la relation entre la forme et le contenu est fondé sur des liens directs; ce qui n'est pas le cas de la SF où il y a nécessairement la médiation des rapports établis entre les constituants (cf. la notion de catène chez H. Frei 1962).

Tout cela nous conduit à la conclusion que la polylexicalité est à la SF ce que la polysémie est à l'unité simple. Se rattache à ce constat un certain nombre de faits :

1- La polylexicalité fournit à la SF une prégnance symbolique plus grande que celle du signe simple ou du signe construit obtenu par dérivation :

Ex : *chien* [ʃjɛ̃] «Mammifère domestique...»
chiot [ʃjo] «Jeune chien»
terre-neuve [tɛRnoev] «Gros chien à tête large..., originaire de...»

Si la prégnance a un degré minimal dans les mots simples, elle traduit la régularité de la langue avec les séries affixées. Mais avec les SF, le signifiant polylexical, en plus du contrat sémiotique primaire qui l'associe au signifié global, se trouve chargé d'une marque symbolique beaucoup plus grande, provenant d'une forte prégnance laissée par la séquence libre employée comme signifiant.

2- C'est cet abandon des constituants libres de leur fonction sémiotique primaire, selon laquelle ils fonctionnent comme des signes ayant des signifiants et des signifiés propres, qui crée une rupture dans le contrat sémiotique, rupture à laquelle s'ajoute un nouveau contrat scellant la nouvelle relation $S_a / S_é$. La rupture explique le retrait des valeurs initiales des constituants; l'ajout de la nouvelle signification fonde l'épaisseur idiomatique.

3- À cette fonction symbolique on peut ajouter une fonction étymologique assumée par le S_a de la SF. Parmi toutes les unités lexicales, les SF sont les unités qui déclinent immédiatement leur origine. Leurs signifiants indiquent directement leur nature discursive, leur construction syntaxique et les items lexicaux qui leur servent d'habillement.

4- L'aboutissement logique de l'accumulation de telles particularités, c'est le renforcement de la fonction mémorielle qui prend appui sur tout le parcours connu par la SF, de la formation jusqu'à la lexicalisation la plus totale. Y interviennent évidemment tous les faits linguistiques pertinents : la structuration sémantique, les mécanismes référentiels, l'intégration des éléments pragmatiques dans le sens de la SF, les contraintes d'emploi, auxquels on peut ajouter toutes les indications extralinguistiques qui s'y rattachent.

Un exemple : *mettre (avoir) la puce à l'oreille* qui a servi de titre au recueil de Duneton (Stock 1978).

C'est une séquence qui, comparée à son équivalent «éveiller l'attention, la méfiance», sa paraphrase lexicographique, se distingue par cette épaisseur due à la surcharge mémorielle. Se profilent derrière ce signifié global :

- un signifiant qui, malgré sa polylexicalité, fonctionne globalement;
- une dénomination oblique par le biais de laquelle des unités abandonnent leur fonction référentielle de départ (il n'est question dans cette séquence ni de *puce* ni d'*oreille*);
- un mécanisme métaphorique qui justifie le rapprochement entre le sens propre et le sens figuré, chargé d'allusions érotiques (cf. Duneton, 58-63, A. Rey et Chantreau, 992-993 et le *G.L.L.F.*), puis une extension de sens qui lui donne une grande généralisation d'emploi;
- un transfert de domaine qui fait qu'on dénomme une réalité au moyen des termes d'une autre réalité (insecte, partie du corps (soucis,...)).

3. LE CULTUREL

Décrire les mécanismes sémantiques en action dans de telles séquences est certes très éclairant mais ne rend pas directement compte de la «langue comme réceptables de croyances communes [...] révé[ant] beaucoup sur l'imaginaire collectif». «Le lexique [ne] porte-[t-il pas] en lui la marque de croyances profondément enracinées ?» (R. Martin, 1987 : 9).

Cette dimension culturelle est assumée par la fonction mémorielle des unités lexicales, qu'elles soient simples ou complexes (cf. les emprunts, les catachrèses, etc.).

Nous pensons que cette fonction du signe linguistique, jusque-là marginalisée, est très importante. Sans être propre aux SF, — les emprunts, les métaphores et les métonymies lexicalisées, les diverses formes de troncation, les dérivés sont autant de formes concernées par ce phénomène —, cette fonction mémorielle est dans les unités polylexicales à la fois systématique, systémique et évidente :

- systématique, parce qu'aucune SF ne peut exister sans trace de mémoire; son signifiant étant par définition chargé de cette fonction. Sa polylexicalité, de par son origine discursive, témoigne du parcours conceptuel suivi par l'unité polylexicale de la séquence libre avant sa soudure en un signifiant global jusqu'à la globalisation sémantique finale (cf. Mejri 1997)

- systémique, en ce sens qu'elle implique toutes les dimensions du système linguistique : toute séquence porte en elle les traces du fonctionnement du système lors de sa formation (cf. les archaïsmes, par ex.), et qu'elle ne se limite pas au lexique; elle implique aussi la syntaxe (cf. la formation des outils syntaxiques : déterminants complexes, locutions conjonctives et prépositives, adverbess...) et la sémantique (les problèmes relatifs à la globalisation sémantique, à la conceptualisation, aux mécanismes tropiques, etc.)

- évidente, parce qu'elle relève, en synchronie, de la compétence des locuteurs : le signifiant de la SF, en tant que donnée de base, rappelle par certains de ses aspects le signifiant onomatopéique; d'où toutes les actualisations (remotivations) possibles prenant

appui sur le seul signifiant (cf. l'ex. de *noyer le poisson* déjà cité), sans le recours à la moindre donnée étymologique.

Le statut d'articulation supplémentaire permet de rendre compte de tous les transferts qui finissent par se fixer dans les SF qui restent relativement peu étudiés bien qu'ils posent énormément de problèmes dans des domaines aussi variés que l'enseignement, la traduction ou le traitement automatique du langage.

Nous avons essayé de montrer ailleurs (Mejri 1977) que la différence entre les langues réside entre autres dans le choix des domaines de transfert, c'est-à-dire dans le choix des signifiants propres à des domaines comme les parties du corps, les animaux ou des activités variées servant à la dénomination d'autres domaines comme les fruits, les légumes, les pierres précieuses, les outils, les attitudes, etc. Les mécanismes sont identiques d'une langue à une autre mais c'est la nature du domaine-cible qui fait la différence.

Le mot *doigt* par exemple entre dans des expressions françaises renvoyant à l'action (*ne rien faire de ses doigts*, *ne pas bouger le petit doigt*), au nombre (*compter sur le bout des doigts*), à la perfection (*connaître sur le bout du doigt*), etc. Dans le dialectal tunisien, il est plutôt question d'inégalité *ضوابعك مش قد قد* [swabʕikmuʃqadqad] litt. «Tes doigts ne sont pas égaux ou de même niveau», de vol *يدور في ضوابعو* [jdawarfiswabʕu:] litt. «tourner les doigts», d'attitude *كان ضبعك عسل ماتلحشوش الكل* [kansubʕikʕsalma:taləhsuʃilkul] litt. «Si ton doigt est de miel, ne le lèche pas entièrement», etc.

Cette charge culturelle trouve aussi son expression dans une sorte de marques positives et négatives dont les constituants sont porteurs et dont la combinaison fournit la charge sémantique finale de la séquence. Cela donne aux expressions une structure qui fonctionne selon une logique élémentaire naturelle, traduisant la vision que la communauté culturelle a des choses.

Exemples :	perdre la vie → «mourir»		
	- + -	⇒	Sans mécanismes
	trouver la mort → «mourir»		tropiques
	+ - -		
	Casser sa pipe → «mourir»		
	- + - -	⇒	Avec mécanismes
	Manger les mauves par les racines → «mourir»		tropiques
	+ + - -		

Le même mécanisme est en oeuvre dans le dialectal tunisien; ce qui varie, c'est le choix des domaines.

On peut dire autant des outils syntaxiques (cf. la description des relations de lieu en français, Vandeloise 1986 et Mejri 1997), des énoncés proverbiaux (Mejri 1997), etc.

Ainsi le culturel n'est-il plus considéré comme une surcharge qui se greffe sur le langage; il est, au contraire, partie intégrante des mécanismes du système. Avec l'usage de la langue, il se dépose régulièrement et crée avec le temps des fixations fondées sur des mécanismes de prototypie et de transferts multiples tropiques ou autres. C'est la raison pour laquelle on a toujours associé langue et identité culturelle, langue et intelligence : en réalité, tout passe par les filtres décrits plus haut.

4. CONCLUSION

Les SF semblent être un exemple qui illustre très bien cette réhabilitation du culturel en en faisant une articulation dont l'étude permettrait entre autres :

- de préciser davantage cette notion de nouvelle segmentation en en décrivant d'une manière plus fine les mécanismes profonds;
- de dégager dans le cadre de la même langue son impact sur la structuration hyperonymique du lexique en distinguant les domaines structurants des domaines structurés;
- d'exploiter toutes ces données dans des études interlinguistiques pour établir les correspondances possibles entre les langues et pour en tirer profit dans la traduction;
- de prévoir dans le traitement informatique cette segmentation supplémentaire qui, moyennant certains critères syntaxiques et sémantiques, peut aider à faciliter les opérations de reconnaissance.

RÉFÉRENCES

- CURAT, H. (1982) : *La locution verbale en français moderne*, Québec, Presses de l'Université Laval.
- GROSS, G. (1996) : *Les expressions en français; noms composés et autres locutions*, Ophrys, 162 p.
- MEJRI, S. (1994) : «Séquences figées et expression de l'intensité», *Cahiers de lexicologie*, n° 65, 1994-2, pp. 111-122.
- MEJRI, S. (1996) : «Binarisme, dualité et séquences figées», *Les formes du sens, Mélanges Robert Martin*, Duculot, pp. 249-256.
- MEJRI, S. (1997a) : «Défigement et jeux de mots», *Études linguistiques*, vol. 3, *Mélanges Abdelkader Méhiri*, pp. 75-92.
- MEJRI, S. (1997b) : *Le figement lexical. Descriptions linguistiques et structuration sémantique*, Publications de la faculté des lettres La Manouba, Tunisie.

OPPOSÉS DE LANGUE, OPPOSÉS DE DISCOURS : LES LAPSUS ANTONYMIQUES

Pierre J. L. ARNAUD

CRTT, Université Lumière Lyon-2, Lyon, France

La mémoire des mots, ce peut être la manière dont ceux-ci sont représentés dans notre esprit de manière à être compris ou bien à être disponibles en production. La mémoire des mots, ce peut aussi être la trace que chaque mot porte des relations qu'il entretient avec les autres mots de la langue. Les deux mémoires sont impliquées entre autres dans un phénomène banal en production, le lapsus, dont une sous-catégorie comprend le lapsus antonymique. Dire le contraire de ce qu'on voulait dire n'a rien d'exceptionnel, et il arrive qu'on se soit simplement mal exprimé, mais dans la plupart des cas, il s'agit d'un lapsus dans lequel un mot-cible est remplacé par son antonyme. Dans ce qui suit, on examine les erreurs antonymiques présentes dans un corpus de lapsus français, dans la perspective d'en tirer des conclusions sur la nature de l'antonymie et sur les mécanismes de production que ces erreurs mettent en évidence.

L'ANTONYMIE EN LINGUISTIQUE

Il existe une abondante littérature lexicologique-sémantique sur l'antonymie, parmi laquelle les chapitres correspondants des traités de Leech (1974), Lyons (1977) et Cruse (1986) font figure de classiques. On se limitera dans ce qui suit à ce qui est nécessaire pour l'analyse des données.

En s'inspirant de Cruse, mais en modifiant sa terminologie (Cruse utilise le terme *opposites* comme hyperonyme, alors qu'on utilise ici *antonymes*, terme qu'il réserve à la catégorie ici dénommée *polaires*), on peut distinguer trois catégories d'antonymes : 1) les polaires, qui sont gradables (chaud/froid), 2) les complémentaires (mort/vivant), et 3) les inverses (maître/élève). L'antonymie n'est pas une catégorie simple, et certaines paires antonymiques semblent plus fortement opposées que d'autres. Ainsi, selon Cruse, *work/play* et *town/country* sont des antonymes relativement faibles en raison du manque d'une échelle unidimensionnelle claire qui sous-tendrait leur opposition. De la même façon, deux antonymes sont d'autant meilleurs que l'opposition épuise une plus grande proportion de leur sens : *giant/dwarf* sont moins bons que *large/small*. *Teal/coffee* et *gas/electricity* sont ressentis comme des antonymes par certains locuteurs, mais dans des contextes où ils représentent un choix binaire. L'antonymie peut donc être considérée comme une relation prototypique, ce qui s'accorde bien avec les propositions de Chaffin (1992) pour qui les relations sémantiques ne doivent pas être considérées comme des primitifs, mais sont en réalité semblables aux concepts en ce qu'on peut y démontrer

l'existence de gradients de typicité, que des comparaisons de similarité sont possibles, et que des phénomènes d'instanciation par le contexte y apparaissent.

En attaquant les problèmes sous l'angle de la linguistique de corpus, Mettinger (1994) a analysé le comportement en discours des antonymes dans un ensemble de romans en anglais. Selon cet auteur, la recherche linguistique antérieure sur les antonymes était fortement théorique et fondée sur une approche déductive et logique tendant à négliger les données réelles. Mettinger distingue quatre catégories d'antonymie : l'adversativité, l'opposition sémantique systémique, l'opposition sémantique non systémique, et le contraste. L'adversativité dépend du monde réel ou mental. L'opposition sémantique systémique, qui repose sur la structure sémantique du lexique, est stable et relativement indépendante du contexte et on peut dire d'une paire d'occurrences qu'il s'agit d'opposés systémiques si on peut trouver un archilexème commun. L'opposition sémantique non systémique correspond à des paires de mots qui ne sont opposés que par l'intermédiaire des connaissances encyclopédiques. L'un des exemples de Mettinger est *lifeliterature*, qui sont opposés dans un passage de critique littéraire. Les *contrastés* apparaissent souvent dans un nombre réduit de cadres syntagmatiques tels que *X ou Y*, *X et Y*.

La linguistique de corpus est également représentée par Fellbaum (1995), qui a examiné les cooccurrences d'antonymes nominaux et verbaux dans le corpus de Brown (Kucera et Francis 1967), comme Justeson et Katz (1991) l'avaient fait pour les adjectifs. Fellbaum montre que les antonymes verbaux et nominaux sont présents dans le corpus en cooccurrence intra-phrasede façon hautement significative. Il apparaît également que des concepts opposés peuvent être coprésents dans des phrases sous la forme de mots de classes différentes. Comme chez Mettinger, un certain nombre de cadres syntaxiques sont mis en évidence comme, pour le nom, *X as well as Y*; *from X to Y*; *now X, now Y*.

L'ANTONYMIE EN PSYCHOLINGUISTIQUE

Une constatation classique (cf. Clark, 1970; Hörmann, 1972) est que dans les expériences d'associations verbales, un adjectif qui a un antonyme direct (*grand/petit*) fait toujours apparaître celui-ci comme réponse primaire, et que ce genre d'association est celui pour lequel les fréquences des réponses primaires sont les plus élevées. Il existerait donc un très fort lien associatif entre antonymes. Dans une représentation décompositionnelle du sens, il n'y aurait qu'un trait de différence entre deux antonymes, qu'une simple inversion de polarité séparerait (on sait qu'en réalité les antonymes prototypiques sont gradables, ce qui pose des problèmes pour une représentation à traits binaires). Dans une perspective associationniste classique, ce lien mental proviendrait des rencontres d'occurrences antonymiques qui seraient souvent voisines en discours (Charles et Miller, 1989), et les liens associatifs existent entre unités lexicales, et non pas entre concepts (Gross et Miller, 1990). Selon ces auteurs, le lexique adjectival, contrairement au lexique des noms, serait structuré par l'antonymie, par des liens directs dans le cas des antonymes directs, et par l'intermédiaire des polaires dans le cas des antonymes indirects (*grand/minuscule*). Charles, Reed et Derryberry (1994) ont obtenu des résultats expérimentaux à l'aide de tâches chronométrées de décision de similarité qu'ils interprètent comme démontrant que l'antonymie directe peut reposer sur des liens associatifs, alors que l'antonymie indirecte et la synonymie reposent essentiellement sur des relations conceptuelles.

Si, on l'a vu plus haut, des recherches ont montré qu'il est exact que les antonymes cooccurrent souvent (Mettinger, 1994; Fellbaum, 1995), et si l'on peut constater que de nombreuses phraséologies comportent des antonymes (*ça ne me fait ni chaud ni froid, sans queue ni tête, petits et grands, été comme hiver, c'est le jour et la nuit*), l'idée que les liens associatifs proviendraient de la cooccurrence en discours a toutefois été critiquée, car on peut suggérer que les antonymes cooccurrent précisément parce qu'ils sont étroitement liés sémantiquement et ainsi inverser l'argument (Levelt, 1989; Murphy et Andrew, 1993).

MODÈLES DE LA LEXICALISATION ET LIEU DE LA SUBSTITUTION

La plupart des modèles psycholinguistiques actuels de la lexicalisation, c'est-à-dire de la production de mots en discours, comportent trois niveaux de représentation des mots, un niveau conceptuel, un niveau d'unités dites *lemmes* qui sont des unités linguistiques répondant à des spécifications conceptuelles et contenant ou renvoyant aux caractéristiques grammaticales des mots, et enfin un niveau de *lexèmes*, terme qui désigne dans ce cadre la représentation phonologique de la forme des mots. Les différences essentielles entre modèles résident dans la présence ou non de rétroaction. Dans les modèles à stades distincts (cf. Levelt *et al.*, 1991), un lemme est activé depuis le niveau conceptuel, et ensuite ce lemme envoie de l'activation vers la couche lexémique. Les lapsus par substitution proviennent d'une compétition entre unités de même niveau. Dans les modèles interactifs ou connexionnistes (cf. Stemmer, 1985; Dell, 1986; Harley, 1993), un ensemble de traits conceptuels envoie de l'activation à des degrés variables vers plusieurs lemmes qui correspondent à tous ou à certains des traits conceptuels; les lemmes activés envoient à leur tour de l'activation à des degrés variables vers des lexèmes en fonction des phonèmes que ces derniers ont en commun. Il se produit une réverbération de l'activation entre niveaux, et les lapsus se produisent lorsqu'une unité-erreur atteint par sommation un niveau d'activation supérieur à l'unité-cible et suffisant pour remporter la compétition.

Les modèles à deux stades séparés ne parviennent pas à rendre compte de l'existence statistiquement significative de lapsus mixtes sémantiques-formels (Harley, 1984), que les modèles interactifs expliquent parfaitement, et c'est à ces derniers qu'on s'intéressera dans ce qui suit. Par ailleurs, les lapsus par substitutions ne peuvent suffire à eux seuls à prouver la nécessité d'une couche lemmatique, malgré des indices qui tendent dans cette direction comme la concordance de genre entre cible et erreur dans des langues comme le français et l'allemand. Si l'on s'en tient à la seule lexicalisation, ce sont en réalité les lapsus d'une catégorie marginale, celle des mélanges sémantiques, qui démontrent la nécessité de lemmes dans un modèle de la production (cf. Arnaud, en préparation). La réflexion sur la nature des lemmes n'a cependant guère avancé depuis leur introduction par Kempen et Huijbers en 1983.

LES DONNÉES

La collection de lapsus français, dont les lapsus antonymiques étudiés ici constituent un sous-ensemble, comporte 2 400 erreurs rassemblées extensivement par l'auteur (pour la méthodologie, cf. Arnaud, 1997). Sur ce total, 1 087 lapsus constituent des substitutions de mots entiers, dont 332 sont des substitutions sémantiques pures ou mixtes, c'est-à-dire des substitutions où mot-cible et mot-erreur présentent une ressemblance de sens subjective. Une substitution mixte manifeste une ressemblance sur

deux ou exceptionnellement trois aspects, sémantique et formel par exemple. Les substitutions mixtes ne seront pas distinguées des pures dans ce qui suit.

Certains lapsus antonymiques ne sont détectables que par la contradiction logique ou encyclopédique qu'ils entraînent :

- (1) (le commentateur explique que, cette année, le Président de la République reste debout pendant tout le défilé du 14 juillet) l'année passée il s'asseyait au passage des drapeaux et des étendards
- (2) a indiqué que la frontière orientale de la Pologne était tout indiquée — c'était la ligne Oder-Neisse
- (3) ce qui est marrant, c'est le scoop Mitterrand est de droite (l'énoncé est ironique et veut se moquer d'un "non-scoop", la conversation tournant autour d'un ouvrage qui "découvre" que Mitterrand était un homme de gauche; la cible était donc *gauche*)

Il est fatal qu'en situation de collecte extensive, certains de ces lapsus les moins évidents passent inaperçus, et les statistiques sur les antonymes devront être considérées comme plus indicatives que définitives (même si, dans l'ensemble, on peut considérer les données sur les substitutions de mots comme fiables, cf. Arnaud, 1997).

Parmi les lapsus sémantiques, 91 concernent des substitutions antonymiques simples, dans lesquelles erreur et cible sont des opposés monolexémiques; ce nombre n'inclut pas 22 erreurs par antonymes de discours (voir plus bas). Le pourcentage de lapsus antonymiques parmi les substitutions sémantiques est de 27,41 %, et est comparable à celui publié par Hotopf (1982) qui était de 31,25 % sur ses données anglaises.

De nombreux lapsus impliquent des antonymes lexicaux des trois catégories classiques mentionnées plus haut.

- (4) objectivement, j'ai tort/j'ai raison
- (5) i faut que je range la vaisselle sale/leuh, propre
- (6) c'est toi qui as fermé/ouvert hier en rentrant ?
- (7) i fait pas froid/li fait pas chaud dans la maison
- (8) et tes cactus, tu les rentres l'été ?
- (9) comment ils parviennent à se mettre en position haut/pardon, en position basse
- (10) nous sommes dans une situation telle que si nous n'avons pas une hausse rapide des taux d'intérêt/une baisse rapide des taux d'intérêt,
- (11) si vous saviez ce que ça me compliquerait/ce que ça me simplifierait l'existence

D'autres substitutions impliquent des antonymes sans doute moins prototypiques, mais qui, *dans le contexte*, sont parfaitement perceptibles comme tels.

- (12) j'ai quelques renseignements à vous donner (< demander)
- (13) c'est vrai que quand on publie trop les tirages baissent/laugmentent, pardon
- (14) ce n'est pas parce que je suis le gendrelque je suis le beau-père, pardon, de M. Pierre Botton,
- (15) un des seuls musiciens /fr/létrangers invités aux Concerts Spirituels à Paris
- (16) fixant la liste des industries qui seront nationalisées/privatisées, pardon

Un autre sous-ensemble comprend principalement des suites N+Adj avec substitution de l'adjectif. Ce qui distingue ces erreurs des précédentes est le fait que les syntagmes impliqués sont probablement préfabriqués (*chunked*) et qu'on peut voir là une substitution de syntagmes et non d'adjectifs seuls. De nombreux cas impliquent notamment *dernier/prochain* :

- (17) c'est pour ça que j'ai pris de la très bonne viande la prochaine fois/non, la dernière fois
- (18) les principes qu'on avait mis au point l'année prochaine

On peut regrouper dans une autre sous-catégorie des substitutions pour lesquelles la paire cible/erreur ne serait probablement pas perçue comme antonymique par des informateurs (la phrase-test *X est l'opposé de Y* apparaît comme anormale ou douteuse, par exemple). Il s'agit des désignations de référents qui sont en contraste dans le discours, voire dans la situation d'énonciation (catégories *contraste* et *adversativité* de Mettinger, 1994).

- (19) je sais que les clavecinistes et les puristes vont s'étonner qu'on puisse jouer Rameau au clavecin

L'information commentée dans ce passage n'aurait rien d'étonnant, mais la discussion porte sur le fait que la personne interviewée est un pianiste qui aime jouer Rameau, et la cible était *piano*. *Piano* et *clavecin* ne sont pas des antonymes, mais des cohyponymes, et c'est la situation d'énonciation qui les met clairement en contraste. De la même façon, *supérieur* et *secondaire* s'opposent dans une réunion d'universitaires où l'on parle d'horaires :

- (20) mon mari est enseignant dans le supérieur/dans le secondaire, je veux dire

Garrett (1992) rapporte des cas semblables, où, par exemple, *deaf* se substitue à *blind* ou *syntactic* à *semantic*. Les membres de ces paires ne sont pas antonymiques, mais l'intention communicative était "sourde seulement, pas aveugle", "seulement syntaxique, donc pas sémantique".

Ce genre de lapsus est particulièrement fréquent entre adjectifs de nationalité.

- (21) on n'aura plus une Université de Lausanne et une Université de Genève, mais une seule université alémanique

Alémanique se substitue à *romane* avec lequel il est en contraste implicite dans le discours, dans lequel il n'a à aucun moment été question explicitement des universités de la Suisse germanophone.

- (22) (il est question des noms en Mac- dont il est parfois impossible de dire s'ils sont écossais ou irlandais; un contre-exemple indubitablement irlandais vient d'être cité) oui, ça c'est écossais
(NB il s'agit bien d'un lapsus, que le locuteur a reconnu comme tel)
- (23) - Vous allez manifester à Rome. Est-ce que ce n'est pas une immixtion dans les affaires françaises ?
- Vous voulez dire italiennes ?
- Italiennes.
- (24) (pendant un épisode de la guerre des pêches entre la France et l'Espagne) le tribunal maritime vient d'imposer une amende très élevée à ce chalutier donc françaisespagnol
- (25) il pense en effet que les Amérlque les Allemands ont fait beaucoup de mal (de Henry Morgenthau, secrétaire américain au trésor en 1944, et de ses projets pour l'Allemagne).

CATEGORIES GRAMMATICALES

Le tableau 1 reproduit les données de catégories grammaticales pour les erreurs sémantiques. La différence entre lapsus sémantiques non antonymiques et lapsus antonymiques est très significative ($\chi^2 = 132,06 - 3 \text{ d.d.l.} - p < 0,001$), ce qui confirme la constatation de Garrett (1992) sur la part dominante de l'antonymie dans les substitutions adjectivales anglaises. Ceci peut constituer un argument en faveur de l'idée que la structure sémantique du lexique adjectival pourrait reposer sur l'antonymie (voir plus haut), mais pas permettre de décider si la substitution est d'origine conceptuelle ou linguistique. En outre, le fait que noms et verbes soient concernés, à un degré certes moindre que les adjectifs, mais néanmoins réel, est à rapprocher des données de Fellbaum (1995).

	non-antonymiques		antonymiques	
	n	%	n	%
Noms	202	87,07	18	19,78
Verbes	20	8,62	17	18,68
Adjectifs	8	3,44	42	46,15
Adverbes	2	0,86	14	15,38
total	232	100	91	100

Tableau 1 : Substitutions sémantiques et catégories grammaticales

LIEU DES SUBSTITUTIONS ANTONYMIQUES

Où se produisent les substitutions antonymiques ? Cette question, notons-le, n'est pas la même que : Comment se produisent les substitutions ? En effet, que le mécanisme soit intra ou inter-niveaux, et quelle que soit la cause ultime qui amène une unité-erreur à un degré d'activation supérieur à celui de l'unité-cible, fluctuations aléatoires autour de la valeur de repos, rémanence d'une activation antérieure, activation provenant de causes centrales autres que le message¹ comme une contamination perceptive, voire causes de type freudien inaccessibles à la démonstration scientifique, la question de la forme et de l'organisation des représentations est capitale.

Comme on l'a vu, les antonymes sont fortement liés associativement, et les substitutions d'antonymes "classiques" (erreurs 4 à 11) pourraient provenir de mécanismes associatifs où, lorsqu'une unité reçoit de l'activation, elle en envoie une forte dose à ses associés les plus proches. Si on accepte que les associations existent entre unités linguistiques (voir plus haut) et non entre concepts, les substitutions se produiraient entre lemmes, voire entre lexèmes, et la proximité sémantique erreur/cible ne serait alors au fond qu'un épiphénomène, puisque si le sens a pu jouer un rôle dans l'établissement de l'association, il n'intervient pas dans le déclenchement de l'erreur. Quelles raisons peuvent permettre de postuler une origine associative aux lapsus ? Leur caractère involontaire, irrépressible, rappelle les phénomènes d'amorçage (McNamara, 1992, 1994). On considère souvent que les effets d'amorçage qui apparaissent dans divers paradigmes expérimentaux sont tellement rapides qu'ils ne permettent pas un accès lexical complet, c'est-à-dire jusqu'à la représentation conceptuelle (Le Ny, 1989), et Shelton et Martin (1992) ont obtenu expérimentalement des effets d'amorçage automatique pour des associés, mais pas pour des mots proches sémantiquement mais non associés. Cependant, des expériences récentes d'amorçage par amorces masquées, donc ne pouvant parvenir à la conscience des sujets, avec tâches de décision lexicale et de dénomination (Perea et Gotor, 1997) ont démontré des effets extrêmement rapides non seulement pour des paires de mots associés, mais aussi pour des paires à lien sémantique mais non associées. Il est donc possible que l'amorçage automatique se produise aussi pour des non-associés, et ceci réduit la nécessité de postuler un lieu linguistique, c'est-à-dire infra-conceptuel, spécialement pour les substitutions d'antonymes classiques.

Les substitutions d'antonymes classiques ne sont pas les seules, on l'a vu, et l'examen des autres catégories est indispensable. *Clavecin* et *piano* (erreur 19) et les adjectifs de nationalité (21 à 25), on l'a vu également, sont des cohyponymes, mais sont opposés par le discours. Les circonstances d'énonciation sont telles qu'il y a une forte opposition, même implicite. Rappelons-nous qu'il n'est à aucun moment question de la Suisse alémanique dans le discours précédant l'apparition de l'erreur 21; par ailleurs, le discours précédant l'erreur 23 ne comportait pas d'occurrence de *France* ou de *français*, et il n'est donc pas possible dans ces deux cas d'invoquer un état d'activation rémanente (au moins à court terme) des unités linguistiques correspondantes. L'explication causale pourrait être que l'opposition de deux concepts les met tous deux en un état d'activation élevé et que, quel que soit le mécanisme, compétition intra-niveau ou résonance inter-niveaux entre unités partageant des traits, le concept-erreur a peu de mal à remporter la compétition et donc à être lexicalisé à la place de l'autre.

¹ Ce terme désigne le contenu de communication pré-verbal.

Il existe d'autres cas de lapsus qui font penser à un phénomène général de permutation de contraires. C'est ainsi qu'on rencontre des oppositions d'antonymes morphologiques, dont on peut se demander s'ils sont lexicaux, c'est-à-dire préfabriqués, ou générés au coup par coup, ou encore si leur représentation n'est pas à mi-chemin entre ces deux cas :

(26) pendant ce temps le petit truc i gèleli s'dégèle

On trouve également des cas d'inversion de polarité non plus lexicale, mais grammaticale :

(27) i faut pas être réalisteli faut être réaliste, pardon

(28) J'veux dire, i z'ont horreur des granulés pour hamsters. Ils les détestent paslils les détestent.

(29) j'suis pas sûre qu't'as bien remué la salade; parce que t'as mis beaucoup trop de temps\pas assez de temps, j'veux dire

On ne peut exclure que de tels lapsus aient leur origine dans un module grammatical, donc linguistique, mais y voir la conséquence d'un changement de polarité conceptuel présente l'avantage d'une explication unique pour les lapsus antonymiques, dont on a vu qu'ils constituent une catégorie remarquable par le nombre et la variété.

Que la gestion mentale des oppositions présente souvent des difficultés peut être illustré par le fait suivant, observé lors du colloque où cet article a été présenté. Une participante, qui faisait une communication sur les nominalisations, expliqua d'abord que, historiquement, certains verbes précédaient les noms correspondants. Puis, projetant un transparent, elle en fit un commentaire qui pouvait sembler aller en sens contraire. Quelques instants plus tard, elle s'interrompit dans son discours, pour finir par se demander à voix haute si elle n'avait pas dit le contraire de ce qu'elle avait voulu dire — à ce stade, l'assistance était de toute façon irrémédiablement perdue ! S'il n'y a pas eu là de lapsus par substitution de mots proprement dit, on n'en a pas moins eu affaire à un phénomène très proche de certains des lapsus étudiés ci-dessus.

Ces considérations ne peuvent que rester spéculatives, car, dans la masse de données que nous fournissent les lapsus substitutionnels, très peu constituent en fait des preuves permettant de trancher entre les détails de plusieurs modèles. Une raison en est que les lapsus naturels ne sont pas susceptibles de manipulations expérimentales. Une autre raison est que la nature des lemmes est encore insuffisamment spécifiée dans les modèles, notamment en ce qui concerne leur contenu sémantique. Or, la question de savoir ce qui du sens des mots est linguistique et ce qui est extra- (supra-) linguistique n'est pas sans intérêt...

RÉFÉRENCES

- ARNAUD, P. J. L. (1997) : “Les ratés de la dénomination individuelle : typologie des lapsus par substitution de mots”, C. Boisson et Ph. Thoiron (dir), *Autour de la dénomination*, Lyon, P.U.L., pp. 307-331.
- ARNAUD, P. J. L. (en préparation) : “Target-error resemblance in word-substitution speech errors and the mental lexicon”.
- CHAFFIN, R. (1992) : “The concept of a semantic relation”, A. Lehrer & E. F. Kittay (eds), *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization*, Hillsdale, N.J., Erlbaum, pp. 253-288.
- CHARLES, W. G. & G. A. MILLER (1989) : “Contexts of antonymous adjectives”, *Applied Psycholinguistics*, 10, pp. 357-375.
- CHARLES, W. G., REED, M. A. & D. DERRYBERRY (1994) : “Conceptual and associative processing in antonymy and synonymy”, *Applied Psycholinguistics*, 15, pp. 329-354.
- CLARK, H. H. (1970) : “Word associations and linguistic theory”, J. Lyons (ed.) *New Horizons in Linguistics*, Harmondsworth, Penguin, pp. 271-286.
- CRUSE, D. A. (1986) : *Lexical Semantics*, Cambridge, C.U.P.
- DELL, G. (1986) : “A spreading-activation theory of retrieval in sentence production”, *Psychological Review*, 93, pp. 283-321.
- FELLBAUM, C. (1995) : “Co-occurrence and antonymy”, *International Journal of Lexicography*, 8, pp. 281-303.
- GARRETT, M. F. (1992) : “Lexical retrieval processes: semantic field effects”, A. Lehrer & E. F. Kittay (eds), *Frames, Fields, and Contents: New Essays in Lexical and Semantic Organization*, Hillsdale, N.J., Erlbaum, pp. 1-15.
- GROSS, D. & K. J. MILLER (1990) : “Adjectives in WordNet”, *International Journal of Lexicography*, 3/4, pp. 265-277.
- HARLEY, T. A. (1984) : “A critique of top-down independent levels models of speech production: evidence from non-plan-internal speech errors”, *Cognitive Science*, 8, pp. 191-219.
- HARLEY, T. A. (1993) : “Phonological activation of semantic competitors during lexical access in speech production”, *Language and Cognitive Processes*, 8, pp. 291-309.
- HÖRMANN, H. (1971) : *Introduction à la psycholinguistique*, Paris, Larousse.

- HOTOPF, W. H. N. (1982) : "Semantic similarity in whole-word slips of the tongue", V. A. Fromkin (ed.), *Errors in Linguistic Performance*, New York, Academic Press, pp. 97-109.
- JUSTESON, J. S. & S. M. KATZ (1991) : "Co-occurrences of antonymous adjectives and their contexts", *Computational Linguistics*, 17, pp. 1-19.
- KEMPEN, G., & P. HUIJBERS (1983) : "The lexicalization process in sentence production and naming: indirect election of words", *Cognition*, 14, pp. 185-209.
- KUCERA, H. & W. N. FRANCIS (1967) : *Computational Analysis of Present-day American English*, Providence, Brown U.P.
- LEECH, G. (1974) : *Semantics*, Harmondsworth, Penguin.
- LE NY, J.-F. (1989) : "Accès au lexique et compréhension du langage : la ligne de démarcation sémantique", *Lexique* (Lille), pp. 65-85.
- LEVELT, W. J. M. (1989) : *Speaking: from Intention to Articulation*, Cambridge (Mass.), M.I.T. Press.
- LEVELT, W. J. M., SCHRIEFERS, H., VORBERG, D., MEYER, A., PECHMANN, T. & J. HAVINGA (1991) : "Normal and Deviant lexical processing: A reply to Dell and O'Seaghdha", *Psychological Review*, 98, pp. 615-618.
- LYONS, J. (1977) : *Semantics, Vol. 1*, Cambridge, C.U.P.
- MCNAMARA, T. P. (1992) : "Theories of priming: I. Associative distance and lag", *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, pp. 1173-1190.
- MCNAMARA, T. P. (1994) : "Theories of priming: II. Types of primes", *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, pp. 507-520.
- METTINGER, A. (1994) : *Aspects of Semantic Opposition in English*, Oxford, Clarendon Press.
- MURPHY, G. L. & J. M. ANDREW (1993) : "The conceptual basis of antonymy and synonymy in adjectives", *Journal of Memory and Language*, 32, pp. 301-319.
- PEREA, M. & A. GOTOR (1997) : "Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming", *Cognition*, 62, pp. 223-240.
- SHELTON, J. R. & R. C. MARTIN (1992) : "How semantic is automatic semantic priming ?", *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, pp. 1191-1210.
- STEMBERGER, J. P. (1985) : "An interactive activation model of language production", A. W. Ellis (ed.), *Progress in the Psychology of Language, Vol. 1*, London, Erlbaum, pp. 143-186.

LES PROVERBES : DES DÉNOMINATIONS D'UN TYPE «TRÈS TRÈS SPÉCIAL»

Georges KLEIBER

Université des sciences humaines, Strasbourg 2 et Scolia, France

La nuit, tous les proverbes sont gris

INTRODUCTION

Un premier constat pour commencer : les proverbes ont actuellement la cote chez les sémanticiens et pragmaticiens. En témoigne une abondante littérature récente¹ qui essaie de cerner la définition et le fonctionnement sémantico-pragmatique des proverbes. Un second constat, toujours pour commencer : cette littérature présente un désaccord profond sur la manière de les envisager. Deux tendances contraires s'y font jour. Une tendance «optimiste» qui pense qu'il est possible de définir linguistiquement le proverbe et une tendance plutôt «pessimiste» qui renoue en quelque sorte, sans s'y arrêter toutefois — heureusement —, avec le défaitisme définitoire d'un A. Taylor (1931)². Les premiers considèrent que les proverbes constituent une catégorie linguistique suffisamment homogène pour être décrite de façon unitaire, avec des propriétés et des tests de reconnaissance linguistiques spécifiques³, alors que les seconds⁴ trouvent qu'une telle attitude est excessive : les proverbes forment une catégorie plutôt hétérogène⁵, à laquelle

¹ Voir depuis notre tentative de 1989 (reprise dans Kleiber, 1994), les articles d'Anscombe (1989, 1990, 1994 et 1995), Arnaud (1991, 1992), Arnaud et Moon (1993), Somolinos Rodriguez (1993), Franken (1995), Michaux (1995 et 1996), Forest (1996) et Gouvard (1996 et à paraître).

² «*The definition of a proverb is too difficult to repay the undertaking*», Taylor (1931 : 3) cité par Schulze-Busacker (1984 : 135).

³ Se laissent placer sous cette bannière les travaux de Kleiber (1989), Anscombe (1994), et Arnaud (1991 et 1992).

⁴ Empressons-nous de souligner (i) que les travaux que nous rangeons dans ce deuxième courant, à savoir essentiellement ceux de Franken (1995), Michaux (1995 et 1996) et de Gouvard (1996 et à paraître), ne se cantonnent nullement dans une position négative stérile, mais présentent des propositions nouvelles stimulantes, notamment celles qui se rattachent à la poéticité du proverbe, et tracent des perspectives destinées à faire avancer positivement les choses et (ii) qu'il ne s'agit évidemment pas d'un jugement de valeur dépréciatif de notre part, bien au contraire.

⁵ «Mieux vaut se faire à l'idée, comme le suggèrent également Rodegem (1984), Michaux (1995) et Franken (1995), que les énoncés que l'on range sous l'étiquette de *proverbes* sont loin

les traits définitoires et les tests mis en relief par les premiers ne s'appliquent qu'imparfaitement.

Je me propose dans ce travail de reprendre le débat et d'examiner en détails, non pas le problème dans son entier, mais une des pièces litigieuses du dossier, celle qui concerne un des plus importants traits définitoires généralement attribué aux proverbes, celui d'être un jugement collectif et non un jugement individuel. La tradition veut que les proverbes ne soient pas la voix d'un particulier, mais véhiculent l'expression de la «sagesse populaire», *vox populi*, ou encore «sagesse des nations». Les partisans de la première tendance ont essayé, de diverses manières et dans des cadres théoriques différents, d'explicitier ce trait et d'en saisir les manifestations linguistiques au travers d'une batterie de tests identificatoires. Leur application, par les tenants du courant opposé, à des proverbes particuliers diversifiés est à la source d'une remise en cause assez radicale — et inattendue — à la fois du trait lui-même et des critères linguistiques s'y rapportant. L'enjeu, on le voit, est de taille, puisqu'il porte sur un aspect central de la conception et du rôle des proverbes : un proverbe peut-il ou non constituer un jugement individuel ?

Notre examen se fera en trois parties. Nous exposerons tout d'abord la position classique du proverbe-jugement collectif en rappelant et en revisitant les principaux critères formels qui viennent l'étayer. La deuxième partie donnera la parole aux détracteurs et présentera les principaux arguments avancés pour limiter ou invalider l'efficacité des critères définitoires évalués. Dans la troisième enfin, nous prendrons parti pour la première position à partir de l'hypothèse définitoire que nous avons fournie en 1989, à savoir que les proverbes sont des dénominations d'un type «très très spécial». En même temps qu'elle apportera une réponse à la question centrale posée, notre analyse permettra, chemin faisant, de mieux délimiter les proverbes par rapport aux phrases génériques d'une part et par rapport aux expressions idiomatiques phrastiques d'autre part. Elle aura l'avantage, en outre, d'expliquer *in fine* — ce qui n'est pas fait dans les traitements habituels⁶ et qui représente donc de ce point de vue-là un progrès certain — l'origine du facteur collectif.

1. LES PROVERBES : DES JUGEMENTS COLLECTIFS

On retrouve à peu près dans toutes les définitions lexicographiques des proverbes qu'il s'agit de vérités ou de jugements qui sont communs à tout un groupe social. Vérités *traditionnelles* ou encore appelées *populaires* (Ollier, 1976), parce qu'elles font partie d'un stock ou «trésor de conseils empiriques accumulés au fil du temps par la sagesse populaire» (Anscombe, 1994 : 99). «*Proverbs in general, souligne Norrick (1985 : 40), are traditional by virtue of their being items of folklore. In this sense, [...], traditionality amounts to common use in a linguistic community or in one of its lectal groups over a period of time, say more than one generation.*»

Deux conséquences linguistiques en découlent :

- celui qui emploie un proverbe n'en est pas l'auteur;

de former une classe homogène qui pourrait recevoir une description linguistique homogène, mais constituent en fait un ensemble de *sous-classes*» (Gouvard, 1996 : 54).

⁶ De nombreux auteurs ont repris le trait de dénomination que nous avons postulé en 1989, sans toujours mesurer à leur juste aune les conséquences qu'il entraîne et les données qu'il permet d'expliquer.

- l'auteur d'un proverbe n'est pas un particulier, mais «quelque chose comme une conscience linguistique collective» (Anscombe, 1994 : 100).

1.1 Le locuteur d'un proverbe n'est pas l'auteur du proverbe

La première mérite d'être explicitée. Généralement, on entend essentiellement par là que le jugement parémique exprimé dans le proverbe ne peut être porté au crédit de celui qui emploie le proverbe. Un locuteur qui dit :

Qui trop embrasse mal étreint

n'est ainsi pas le responsable du contenu exprimé par le proverbe. Autrement dit, ce n'est pas lui qui «pense» ou qui est d'avis que si l'on en fait de trop on risque de mal le faire, même si c'est lui qui énonce effectivement le proverbe en question. En termes de polyphonie, s'il est le *locuteur* du proverbe, il n'est pas, par contre, «l'énonciateur du principe qui y est attaché» (Anscombe, 1994 : 100). La vérité générale exprimée par le proverbe a une autre source ou autre voix que celle du locuteur qui l'emploie. Dans le cadre de la théorie de la pertinence (Sperber et Wilson, 1989), on parle d'énoncé *échoïque* pour rendre compte du fait que le locuteur produit un énoncé qui n'est qu'un écho de propos ou de pensées d'autrui⁷.

Ce qui est souvent ignoré ou reste implicite, même si les manifestations de ce phénomène sont signalées ailleurs, au niveau de la fixité ou rigidité relative du proverbe, c'est qu'il y a un deuxième sens à dire que le locuteur n'est pas l'auteur du proverbe. Il n'est pas non plus le responsable de la forme du proverbe, c'est-à-dire du choix des mots, de leur combinaison, du processus métaphorique choisi s'il y en a un, etc. Ni le traitement polyphonique ni l'analyse en termes d'énoncé échoïque ne rendent explicitement compte de cet aspect-là des choses, puisqu'ils n'exigent pas qu'il y ait eu énoncé antécédent effectif et donc répétition fidèle de paroles prononcées. Il faut uniquement que ce soit la pensée ou le contenu qui ne soit pas propre au locuteur, mais provienne d'autrui. On en tient une preuve toute relative dans la perte du caractère proverbial lorsqu'on paraphrase un proverbe, c'est-à-dire lorsqu'on explicite son sens. Quoiqu'on continue de respecter le contenu du proverbe ainsi explicité et que la responsabilité du locuteur ne se trouve donc pas engagée vis-à-vis du principe exprimé, puisque celui-ci reste grosso modo le même que celui du proverbe, l'expression produite perd dans l'opération son statut de proverbe. C'est ainsi que si on paraphrase :

*La langue va où la dent fait mal
À chaque pot son couvercle⁸
Loin des yeux, loin du coeur*

par des énoncés explicatifs tels que :

⁷ De façon plus précise, comme le développe Gouvard (1996 : 56-57) en citant Sperber et Wilson (1989 : 357), un énoncé est échoïque lorsqu'on l'interprète de façon échoïque, c'est-à-dire lorsqu'on lui attribue une interprétation qui «doit sa pertinence au fait que le locuteur se fait à sa façon l'écho des propos ou des pensées d'autrui», dans le but de véhiculer une information qui lui est propre.

⁸ Proverbes cités par Gouvard (1996).

*On parle plus volontiers de ses peines
Chaque femme ou chaque homme finit par trouver l'homme ou la femme qui lui
convient
L'éloignement de deux êtres qui s'aiment fait diminuer leur amour*

on observe que ceux-ci y ont laissé des plumes ... proverbiales. Surtout les deux derniers, parce que le premier, étant donné précisément sa forme, peut encore passer pour un *écho* de forme et de contenu, c'est-à-dire pour un proverbe.

Il n'y a rien d'étonnant, étant donné l'oubli de la forme dont nous venons de parler, que les tests destinés à illustrer que le locuteur d'un proverbe n'est pas l'auteur du proverbe concernent avant tout le contenu notionnel du proverbe. Anscombe (1994 : 100), s'appuyant sur l'étude de Ducrot (1975) sur *je trouve que*, note que les proverbes ne se combinent guère avec cette expression performative d'opinion individuelle, lorsqu'il «s'agit d'exprimer une adhésion générale au principe exprimé par le proverbe» :

**Je trouve que la fortune sourit aux audacieux
*Je trouve que qui va à la chasse perd sa place
*Je trouve que petite pluie abat grand vent
Je trouve que prudence est mère de sûreté⁹

Pour les mêmes raisons, il est difficile d'avoir les combinaisons *je crois que + proverbe*, *je pense que + proverbe*, *selon moi + proverbe*, *à mon avis + proverbe*, etc. (Gouvard, 1996 et Anscombe, 1994). Elles posent toutes l'identité entre le locuteur et l'énonciateur de l'opinion, alors que le proverbe postule au contraire la dissociation de ces deux rôles :

*?Je crois que qui trop embrasse mal étreint
?Je pense que qui trop embrasse mal étreint
?Selon moi, qui trop embrasse mal étreint
?À mon avis, qui trop embrasse mal étreint*

Les adverbes d'énonciation *franchement* et *visiblement* donnent lieu au même constat (Anscombe, 1994 : 101) :

*? Visiblement, qui ne risque rien n'a rien
??Franchement, pas de nouvelles bonnes nouvelles*

Le cas d'*estimer* est traité différemment. Comme il «admet la reprise d'un jugement dont le locuteur n'est pas l'auteur» (Anscombe, 1994 : 100), on s'attend à ce que les proverbes puissent lui servir de complément. Et pour Anscombe le résultat est effectivement meilleur, même s'il reconnaît dans une note et par les points d'interrogation dont il préfixe les deux premiers exemples de ci-dessous que la combinaison est généralement jugée imparfaite¹⁰ :

*?J'estime que la fortune sourit aux audacieux
?J'estime que qui va à la chasse perd sa place*

⁹ La stellarisation est celle d'Anscombe.

¹⁰ D'un autre côté, il souligne qu'elle est «parfois même acceptée sans problème par les sujets parlants» (1994 : 100). Nous reviendrons sur ce point dans notre dernière partie.

*J'estime que prudence est mère de sûreté
J'estime que le soleil luit pour tout le monde*

Il faut bien souligner que le test de *je trouve que* + *proverbe* n'est pertinent que si la portée du jugement s'exerce sur la validité du contenu général exprimé par le proverbe. Si le locuteur n'est pas l'auteur du proverbe, il est par contre «l'auteur» de son emploi : c'est lui qui est responsable de l'énonciation du proverbe et du principe qui y est attaché, et qui, selon le type d'emplois¹¹, «endosse la responsabilité de déclarer ce principe applicable *hic et nunc*» (Anscombe, 1994 : 100). Si donc les performatifs d'opinion du type *je trouve que* et autres adverbes d'énonciation incompatibles en portée générale avec le proverbe ont pour objet l'application du proverbe à une situation particulière, l'interdiction d'une combinaison avec un proverbe se fait moins forte, comme le montre Anscombe (1994 : 101) avec les exemples suivants :

*Je trouve que, pour une fois, à quelque chose malheur est bon
Je trouve que, au vu des circonstances, pas de nouvelles, bonnes nouvelles
Je trouve que, dans le cas qui nous occupe, le vin étant tiré, il faut le boire
Visiblement, de nos jours, qui ne risque rien, n'a rien
Franchement, au vu des circonstances, pas de nouvelles, bonnes nouvelles
À mon avis, dans ton cas, prudence est mère de sûreté*

auxquels on peut ajouter :

*Je crois que, avec Fred, loin des yeux, loin du coeur
Je pense que, dans cette affaire, qui trop embrasse mal étreint
Selon moi, étant donné la situation, prudence est mère de sûreté*

Dans ce cas, le jugement exprimé par le locuteur-énonciateur ne dit pas que ce qu'exprime le proverbe est vrai, c'est-à-dire que le contenu notionnel est valide en général, mais que le proverbe est vrai dans telle ou telle situation particulière envisagée, c'est-à-dire que le proverbe s'applique à ou se trouve vérifié par telle ou telle situation particulière envisagée.

1.2 Jugement individuel et jugement collectif

Les tests employés jusqu'à présent permettent uniquement de conclure que ce n'est pas le locuteur d'un proverbe qui en est l'auteur. Ils n'autorisent pas, du moins directement, à affirmer qu'il ne s'agit pas d'un jugement individuel¹². L'énonciateur, c'est-à-dire le responsable du proverbe, pourrait en effet être un autre particulier. La plupart des commentateurs rejettent cette possibilité : un particulier ne peut être tenu pour responsable d'un proverbe. Celui-ci est le fait d'un énonciateur collectif, de la *vox populi* et non de tel ou tel individu particulier. La polyphonie inhérente au proverbe met donc aux prises un particulier, le locuteur, qui énonce le proverbe, et un énonciateur collectif (cf. supra la *conscience linguistique collective* d'Anscombe). L'énoncé échoïque que représente le proverbe «n'est attribuable à aucune source précise, mais au peuple dans son ensemble»

¹¹ Nous ne pouvons dans le cadre de ce travail développer ce point comme il faudrait. Signalons simplement qu'un locuteur peut employer un proverbe de façon «non spécifique».

¹² Gouvard (1996) est un des rares à faire soigneusement cette distinction.

(Sperber et Wilson, 1989 : 358)¹³, ce qui conduit Gouvard à une redéfinition «écho-écot» du proverbe comme étant en somme *l'écho d'un écho* :

«Un énoncé proverbial est un énoncé dont l'interprétation échoïque implique nécessairement que l'énoncé dont le locuteur se fait l'écho n'est lui-même interprétable que sous une forme échoïque.» (Gouvard, 1996 : 57)

Ce trait permet classiquement de séparer les proverbes des sentences, maximes, aphorismes, apophtegmes, slogans, etc., c'est-à-dire de tous les énoncés sentencieux qui ont un «père» particulier (identifiable ou non). On signale tout aussi traditionnellement que de tels énoncés individuels peuvent devenir proverbes¹⁴. Ce n'est pas parce que l'on connaît l'origine de tel ou tel proverbe particulier que le proverbe en perd pour autant son statut de vérité «collective». La connaissance diachronique que les proverbes :

*À petite cloche, grand son
Honni soit qui mal y pense*

son à l'origine les devises de la maison de Grandson et de l'ordre de la Jarretière¹⁵ ne les empêche pas synchroniquement d'apparaître comme des jugements non plus individuels, mais collectifs, des vérités générales faisant partie du stock de principes généraux communs à toute une communauté (ou *peuple*, pour reprendre le terme de Sperber et Wilson).

De la même manière, la confection de proverbes, activité poético-ludique assez répandue, n'entre pas en contradiction avec le trait de vérité collective attribué au proverbe. Ces proverbes «fabriqués» ou *formes proverbiales*, comme les appelle Gouvard (1996) pour les séparer des proverbes attestés, sont certes des créations individuelles, mais leur «générateur» les a fabriqués de telle sorte qu'ils passent pour ne pas avoir d'auteur particulier, mais relèvent de la voix anonyme collective des proverbes. S'il n'en allait pas ainsi, leur auteur aurait manqué son coup : on ne les reconnaîtrait pas pour ce qu'il veut qu'on les reconnaisse, à savoir des proverbes. Sentences de particuliers passés en proverbes comme néologismes proverbiaux fabriqués par des particuliers ne peuvent ainsi remettre en cause la portée de jugement collectif reconnu aux proverbes. Dans les deux cas, celui des phrases devenues proverbes comme celui des formes proverbiales, il faut évidemment que l'expression en question réponde aux autres attributs qui définissent le proverbe ou qui, du moins, font que l'on reconnaisse telle ou telle forme comme étant ou comme pouvant être un proverbe.

Le critère classique qu'on associe à la propriété de jugement collectif est l'expression métalinguistique à dire indéfini *comme on dit* avec un *on* révélateur (Kleiber, 1989a et 1994; Anscombe et Gouvard, 1996), qui s'oppose à l'expression à dire défini *comme (le) dit X* où X représente un individu particulier :

Et, comme on dit, qui trop embrasse mal étreint

¹³ Que nous citons d'après Gouvard (1996 : 57).

¹⁴ Faits cités par Gouvard (1996 : 57).

¹⁵ Cf. l'expression *passer en proverbe*.

Comme on dit, jamais à bon chien il ne vient un os (cité par Gouvard, 1996)

Comme dit Rabelais, fais ce que voudras

Comme le dit La Rochefoucauld, le refus des louanges est un désir d'être loué deux fois (cité par Anscombe, 1994)

Anscombe (1994 : 99) cite également des tournures telles que *On a bien raison de dire que ..., si j'en crois la sagesse populaire (var. la sagesse des nations) :*

On a bien raison de dire qu'on n'est jamais trahi que par les siens

Si j'en crois la sagesse populaire / la sagesse des nations, il ne faut jamais mettre la charrue avec les boeufs

On peut aussi citer ici la tournure *comme (le) dit le proverbe :*

Comme le dit le proverbe, un homme averti en vaut deux

dont la particularité, par rapport aux expressions métonymiques du type *la phrase dit / cette expression dit ...,* avec lesquelles on peut être tenté de l'assimiler, consiste à ce que la définitude du proverbe se trouve justifiée par l'apparition même du proverbe comme complément. On ne peut avoir dans les mêmes conditions :

Comme le dit la phrase / l'expression, cherche ton ciel et tu trouveras ton soleil

un tel énoncé nécessitant une justification préalable (c'est-à-dire extérieure à la mention qui suit) de la définitude de la phrase ou de l'expression. L'emploi de l'indéfini confirme cette différence. Face à :

Comme dit un proverbe, le bon est bon, mais le meilleur est meilleur

on ne peut énoncer :

?Comme dit une phrase, les menteries ont courtes jambes

?Comme dit une phrase, les proverbes doivent être décortiqués

sans provoquer une surprise légitime, même si la phrase qui suit est une ... phrase, proverbe ou non.

Le critère du *Comme dit X vs comme on dit* demande toutefois des précisions supplémentaires. Premièrement, il n'est pas exclu d'avoir *comme (le) dit X* avec un proverbe, ainsi que le signale fort justement Gouvard (1996 : 50) :

Comme (le) dit ma concierge, les mouches se reposent toujours sur les chevaux maigres

Comme (le) dit ma grand-mère, vieux mari et jeune femme, des cornes en campagne

mais on observera qu'un tel X ne passe pas pour être l'auteur ou le responsable du proverbe, mais plutôt un utilisateur ou un «habitué» du proverbe. Ma grand-mère ne se trouve pas présentée comme étant à l'origine du proverbe *petite pluie abat grand vent*, mais

le fait de l'employer fréquemment constitue une habitude qui la caractérise¹⁶. On le vérifiera en transformant l'habitualité du dire de *comme dit X* en épisodicité, c'est-à-dire en dire-occurrence (Kleiber, 1987). Les énoncés obtenus :

?*Comme a dit ma concierge, les mouches se reposent toujours sur les chevaux maigres*

?*Comme a dit ma grand-mère, vieux mari et jeune femme, des cornes en campagne*

passent mal la rampe, dans l'interprétation où X est le responsable du proverbe.

Cette mise au point n'est pas sans conséquence pour le critère de la tournure *comme on dit* elle-même. Elle montre en effet que cette tournure n'est nullement une preuve directe du caractère collectif du jugement exprimé par le proverbe, mais représente une manifestation de l'utilisation commune ou partagée par les locuteurs qui en est faite. Elle rejoint un autre trait du proverbe, celui de la collocation avec l'adjectif *connu* (cf. *un proverbe (bien) connu*). Nous aurons l'occasion d'y revenir ci-dessous lorsque nous aborderons la question de la dénomination. Pour le moment, on remarquera en faveur de notre rectification la similitude aspectuo-temporelle avec la structure *comme dit X* : de même que l'on ne peut avoir en interprétation épisodique *comme a dit X + Proverbe*, de même on ne peut avoir en lecture événementielle *comme on a dit + Proverbe* :

?*Comme on a dit, petite pluie abat grand vent.*

Il est significatif de constater que la même contrainte vaut pour l'expression *comme dit le proverbe*¹⁷ :

?*Comme l'a dit le proverbe, rien ne sert de courir il faut partir à point.*

Ce n'est pas pour autant qu'il faut écarter ce critère : l'indéfinitude collective du *on* jointe à l'habitualité du dire permettent de conclure indirectement que son énonciation, lorsqu'elle a lieu, n'est plus le fait d'un particulier, puisque c'est déjà une habitude de le dire (*comme on dit ...*) ou puisque *le proverbe le dit* déjà. D'autre part, il a une vertu supplémentaire, c'est celle de mettre l'accent, par l'intermédiaire du verbe *dire* du *comme*, sur le côté formel de la chose — non touché, on le rappelle, par les verbes d'opinion — en exigeant que la forme ait, d'une manière ou d'une autre, un caractère remarquable, saillant¹⁸

¹⁶ Nous nous séparons sur ce point un tout petit peu de Gouvard (1996 : 50) qui pense que l'individu particulier X est supposé véhiculer la sagesse exprimée par le proverbe, «parce que ce trait fait partie de sa représentation prototypique».

¹⁷ Petite différence entre *comme on dit + proverbe* et *comme dit le proverbe + proverbe*, on peut avoir pour la seconde tournure le pronom *le* en pronom anticipant : ?*comme on le dit, petite pluie abat grand vent* vs *comme le dit le proverbe, petite pluie abat grand vent*.

¹⁸ Ce fait n'est pas particulier au proverbe. Il intervient aussi avec les sentences et autres expressions remarquables (voir aussi pour les expressions idiomatiques). Sans justifications contextuelles spéciales, toujours possibles pour assurer la saillance du dit, on a difficilement *Comme le dit* ou *Comme l'a dit* *La Rochefoucauld / Comme le dit/ l'a dit ma grand-mère, il fait froid le soir/ je mange du veau / j'ai vu passer deux chevaux*, etc.

Le critère du *comme on dit* + *proverbe* n'est donc pas totalement injustifié, mais, pour montrer qu'un proverbe n'est pas un jugement individuel, on aurait dû, en bonne logique, d'abord reprendre le test utilisé pour démontrer que le locuteur n'est pas l'auteur du proverbe en plaçant un autre individu que le locuteur comme le sujet du verbe d'opinion¹⁹. Si des tournures telles que *Paul trouve que / croit que / pense que* + *proverbe*, *Selon Paul*, + *proverbe* apparaissent mal formées, alors on peut logiquement en conclure — directement — que le proverbe n'est pas le fait d'un individu particulier et alors en tirer la conclusion qu'il ne peut s'agir que d'un jugement déjà acquis, qui appartient à une conscience collective, etc. L'essai semble probant, puisque des énoncés tels que :

?*Paul trouve que qui va à la chasse perd sa place*
?*Paul croit que qui trop embrasse mal étreint*
?*Paul pense que qui trop embrasse mal étreint*
?*Selon Paul, qui trop embrasse mal étreint*

suscitent une réaction de surprise, si l'on reconnaît à la phrase complément le statut de proverbe, le proverbe n'acceptant guère d'être remis ou replacé en somme sous la responsabilité d'un particulier. S'il s'agit d'une application particulière, les choses s'améliorent, pour les raisons signalées ci-dessus :

Paul trouve que, pour une fois, à quelque chose malheur est bon
Paul pense que, pour une fois, à quelque chose malheur est bon
Paul croit que, dans cette affaire, qui trop embrasse mal étreint
Selon Paul, étant donné la situation, prudence est mère de sûreté

mais le phénomène du discours oblique fait que ce n'est pas Paul qui est obligé d'avoir prononcé le proverbe, le locuteur pouvant en l'occurrence l'appliquer pour caractériser l'opinion de Paul. Cette possibilité n'a toutefois aucune conséquence sur le résultat que l'on peut tirer du critère : les proverbes apparaissent comme étant des jugements non individuels.

2. UNE REMISE EN CAUSE DU CARACTÈRE «NON INDIVIDUEL» DÉFINITOIRE DES PROVERBES

Ce résultat définitoire se trouve par contre sérieusement ébranlé si l'on arrive à montrer que le critère n'est pas totalement fiable et qu'il existe malgré tout des énoncés où le proverbe se présente comme étant le jugement émis par un individu. Michaux (1995 et 1996) montre tout d'abord qu'on peut avoir deux types de lecture applicative du proverbe, qui ne remettent pas en cause le caractère non individuel du proverbe. À côté de l'application particulière évoquée ci-dessus, qu'elle appelle lecture *métalinguistique locale*, comme en :

Je trouve que, pour une fois, à quelque chose malheur est bon

il est possible d'avoir encore une autre lecture applicative métalinguistique, une lecture *métalinguistique générale*, qui «porte sur l'opinion du locuteur quant à la validité en

¹⁹ Si on a recours à *je trouve que*, c'est à cause de la spécificité de *je trouve que* lorsqu'il est performatif (Ducrot, 1975).

général du proverbe subordonné» et qui donc «projette la validité du proverbe sur l'ensemble des situations vérifiées par le principe général sous-jacent à ce proverbe» (Michaux, 1996). Un locuteur peut ainsi dire :

C'est bien dommage que chien qui aboie ne mord jamais

pour rendre compte «de sa tristesse de constater que dans la vie (c'est-à-dire telle qu'il la connaît), le topos correspondant à la parémie en question est généralement valide» (Michaux, 1996). Il en va de même dans :

Je crains fort que dans l'absolu les cordonniers sont les plus mal chaussés

où il signifie sa crainte de voir le proverbe être vrai «dans l'absolu». Dans ces deux cas, comme dans la situation d'application particulière, il y a une dissociation polyphonique entre locuteur et énonciateur : «le locuteur indique clairement qu'il n'est pas l'auteur du proverbe, même s'il trouve applicable *hic et nunc* (lecture métalinguistique locale) ou en général (lecture métalinguistique générale) le principe qui lui est attaché» (Michaux, 1996). Ce ne sont donc pas de telles interprétations qui peuvent être invoquées pour remettre en cause l'incompatibilité du proverbe avec les jugements individuels. Décisive, par contre, s'avère la combinaison dans laquelle le locuteur exprime son avis sur le contenu et, en même temps qu'il en est le locuteur, se fait donc l'auteur de la parémie exposée :

*Je trouve qu'abondance de biens ne nuit pas
Je trouve que l'argent ne fait pas le bonheur.*

À côté de l'interprétation métalinguistique toujours possible, ces deux énoncés donnent lieu à une lecture portant sur le contenu de la complétive : «le locuteur n'indique pas qu'il considère que le proverbe est un principe valide en général. Il reprend à son compte l'attribution d'un prédicat à un syntagme nominal» (Michaux, 1996).

Gouvard (1995 : 50) souligne également la possibilité d'avoir des constructions qui impliquent la responsabilité énonciative du locuteur :

*Je trouve que l'habit ne fait pas le moine
Je crois qu'il vaut mieux prévenir que guérir
Je crois que l'excès en tout est un défaut*

mais il n'en conclut pas que les proverbes peuvent être des jugements individuels. Il rejette uniquement comme test formel d'identification le critère de l'incompatibilité des proverbes avec les verbes d'opinion individuelle. Il conserve en effet le test de *comme on dit* et maintient dans son approche définitoire du proverbe que le locuteur n'est pas l'auteur du proverbe, puisque celui-ci se trouve défini comme étant l'écho d'un écho.

Michaux (1996), par contre, va jusqu'au bout : les combinaisons *je trouve que + proverbe* à lecture de contenu prouvent, selon elle, qu'un «proverbe peut, dans certaines conditions, être assimilé à un jugement individuel émis par le locuteur hors de toute

situation particulière»²⁰. Et elle formule une explication intéressante pour rendre compte du fait que certains proverbes peuvent accepter la tournure *je trouve que* en lecture subjective et d'autres non. Les proverbes qui intrinsèquement répondent au critère sémantique de prédication originelle²¹ exigé par *je trouve que* acceptent sans trop de peine de devenir la complétive de *je trouve que*. Si, par exemple :

Je trouve qu'il n'est point de sot métier

passe mieux la rampe que :

?*Je trouve que l'eau va à la rivière*

c'est parce que le jugement *Il n'est point de sot métier* est plus facilement un jugement formé sur une prédication originelle, cette base de références subjective déjà disponible, que l'assertion *L'eau va à la rivière*. Pour sauver *Je trouve que l'eau va à la rivière*, il ne reste plus que la lecture métalinguistique générale (ou locale pour *Je trouve que, pour une fois, l'eau va à la rivière*), qui justifie la présence de *je trouve que*.

Le même fonctionnement se laisse appliquer, comme le montre Michaux (1995 et 1996) aux autres verbes d'opinion et aux autres verbes recteurs d'une complétive, à la seule différence qu'il faut à chaque fois que le proverbe présente intrinsèquement les traits exigés par chaque type de verbe. Ainsi *J'estime que + proverbe, je crois que + proverbe, je pense que + proverbe*, etc., seront des combinaisons permises en interprétation non métalinguistique si l'assertion proverbiale répond aux conditions spécifiques exigées respectivement par *j'estime que, je crois que, je pense que*, etc. Cette façon de procéder permet à Michaux (1996) d'expliquer autrement qu'Anscombe pourquoi l'énoncé :

**Je trouve que prudence est mère de sûreté*

est mal formé en lecture de jugement sur le contenu du proverbe, alors qu'une semblable lecture ne pose pas de difficulté pour :

J'estime que prudence est mère de sûreté.

Elle a pour conséquence supplémentaire de conduire à une subdivision des proverbes en deux groupes, selon qu'ils peuvent donner lieu ou non à une lecture sur leur contenu parémique. Le premier groupe, que Michaux (1996) rapproche des phrases génériques typifiantes locales telles que :

Les chats sont affectueux

²⁰ Elle est pourtant à un moment donné de l'article d'un avis différent, lorsqu'elle souligne que l'affirmation d'Anscombe que «le proverbe n'est en rien une opinion personnelle» constitue un fait qui n'est pas critiquable.

²¹ Nous ne discuterons pas de cette notion ici. Voir à ce sujet Ducrot (1975) et son analyse de *Je trouve que ...* et les commentaires et discussions de Michaux (1996). Disons simplement qu'il s'agit de rendre compte du fait que l'opinion exprimée par le locuteur avec *je trouve que* s'établit à partir de données préalables ou *prédication originelle* dont dispose déjà le locuteur.

rassemble «les proverbes dont la nature même peut conduire à un jugement individuel sur leur contenu». Le second, qui est à rapprocher des phrases génériques typifiantes *a priori* comme :

Les castors construisent des barrages

réunit les proverbes «qui, lorsqu'ils permettent l'accès à un jugement d'évaluation, ne peuvent le faire que *via* une interprétation métalinguistique». Résultat final : on ne peut plus définir les proverbes comme des jugements non individuels, puisque le premier groupe se définit précisément par la possibilité de donner lieu dans certaines situations à un jugement individuel.

La chose est gênante, puisqu'on se prive d'un des traits définitoires qui semblait le plus solide, mais elle est logique : si les proverbes peuvent devenir les compléments de verbes d'opinion personnelle, on ne peut plus les définir comme étant des jugements collectifs, des assertions qui ne sont pas des avis personnels, mais qui constituent des vérités faisant partie d'une communauté, appartenant à la «conscience linguistique collective». Les deux conséquences linguistiques qui découlent du caractère *traditionnel* ou *populaire* assigné habituellement aux proverbes, que nous avons mis en avant au début de ce travail, sont évidemment caduques aussi :

- celui qui emploie un proverbe peut en être également l'énonciateur ou l'auteur;
- l'auteur d'un proverbe peut être un particulier.

3. VERS UNE SOLUTION EN TERMES DE DÉNOMINATION

Faut-il aller aussi loin ? Nous ne le pensons pas, malgré les données linguistiques relevées en faveur d'une telle conclusion par Michaux et Gouvard. Il ne s'agit pas de nier ces combinaisons avec des verbes qui engagent la subjectivité du locuteur. On peut effectivement avoir des énoncés tels que :

Je trouve qu'abondance de biens ne nuit pas
Je trouve que l'argent ne fait pas le bonheur
Je trouve que l'habit ne fait pas le moine
Je crois qu'il vaut mieux prévenir que guérir
Je crois que l'excès en tout est un défaut
J'estime que prudence est mère de sûreté
Paul croit qu'il vaut mieux prévenir que guérir
Paul estime que l'argent ne fait pas le bonheur

où c'est le contenu prédicatif SN-SV qui se trouve à chaque fois mis en jeu par le locuteur. Mais on n'est pas pour autant forcé à abandonner l'idée que les proverbes ne sont pas des jugements individuels. Tout simplement, parce qu'une telle conclusion n'est valide que si et seulement si on considère que les prédications en question continuent d'être des proverbes. Or, une telle position n'est nullement obligatoire. Il n'est pas du tout nécessaire de souscrire au maintien du statut de proverbe lorsqu'un proverbe se trouve inséré dans de telles combinaisons.

Notre hypothèse est que dans ces énoncés nous n'avons en fait pas la combinaison d'un verbe d'opinion et d'un proverbe. Ce n'est pas *je trouve que + proverbe*, *Je crois que + proverbe*, *j'estime que + proverbe*, etc., mais *je trouve que + proposition*, *je crois que + proposition*, *j'estime que + proposition*, etc. Comment cela est-il possible, alors qu'il n'y a pas changement de forme entre le proverbe testé et l'énoncé avec le verbe d'opinion ? C'est bien cette identité de forme qui conduit à penser qu'un proverbe peut donner lieu à un jugement individuel. En fait, c'est bien un proverbe qui se trouve pris lorsqu'on applique le test de l'insertion comme complément d'un verbe d'opinion individuelle, mais cette insertion, si elle ne touche pas à sa forme, affecte son statut de proverbe. Ce qui se trouve changé lors de ce placement sous la responsabilité d'un particulier, c'est le caractère de dénomination du proverbe.

3.1 Des dénominations d'un type «très très spécial»

Il nous faut donc au préalable rappeler que les proverbes sont des dénominations d'un type «très très spécial». La particularité fondamentale d'un proverbe est d'être, comme nous l'avons montré ailleurs (1989a et 1994), à la fois, une dénomination²², c'est-à-dire une unité codée, faisant partie du code linguistique, en ce qu'elle nomme une entité générale et non un particulier, et une phrase. Ce double aspect, antinomique, fonde son originalité sémiotique, que nous avons soulignée comme suit : «En tant que phrase, il ne devrait pas être signe (ou unité codée), puisque l'interprétation d'une phrase est une construction et non un donné préalable. En tant que dénomination, il est néanmoins une unité codée, c'est-à-dire un signe. Un signe-phrase donc, qui possède les vertus du signe sans perdre pour autant son caractère de phrase, de même que substantifs, verbes, adjectifs, etc., sont des dénominations qui conservent les attributs spécifiques des catégories grammaticales qu'ils représentent» (1994 : 214).

On notera un premier avantage de cette caractérisation en relation avec notre sujet : le statut de dénomination phrastique des proverbes permet d'expliquer directement pourquoi ils passent pour être des jugements collectifs et non individuels : c'est parce qu'ils sont des phrases-dénominations que la prédication exprimée ne peut être portée au crédit d'un particulier. Cette prédication fait partie du code linguistique commun et, en tant que telle, est une unité dont l'existence n'a plus à être postulée. Elle s'impose à tout locuteur comme toutes les autres dénominations. Observons que cela ne signifie pas qu'il doive lui croire personnellement à son contenu, mais uniquement qu'il est obligé d'accepter que ce contenu, c'est-à-dire le principe attaché au proverbe, est le sens conventionnellement assigné à la phrase proverbiale. De même que le sens d'une unité lexicale est un sens «collectif», de même le contenu, c'est-à-dire la prédication ou le jugement d'un proverbe, est «collectif» et non une affaire de particulier. Le caractère de *vox populi* des proverbes n'est donc qu'une conséquence de leur caractère de dénominations phrastiques.

Ce qui est remarquable, c'est qu'on arrive à rendre compte par la même occasion du fait que le locuteur d'un proverbe n'est pas non plus l'auteur de la forme du proverbe. Nous avons en effet souligné ci-dessus que la plupart des commentateurs oublièrent qu'il y avait un deuxième sens à affirmer qu'un locuteur n'est pas le responsable du proverbe qu'il utilise : il n'est pas non plus maître du choix des mots, de leur combinaison, du processus

²² Les arguments sont nombreux (Kleiber, 1989 et 1994) : (relative) fixité de la forme, apprentissage, présence dans les dictionnaires, (relative) opacité, etc.

métaphorique choisi, s'il y en a un, etc. L'accent est uniquement mis sur le contenu, que ce soit dans l'approche polyphonique ou dans le traitement échoïque. Or, le locuteur d'un proverbe n'est nullement maître de la forme du proverbe qu'il énonce et s'il en va ainsi, c'est bien parce que la phrase énoncée est une dénomination, c'est-à-dire l'association stable d'une forme et d'un sens. De même que l'usager d'un substantif ou d'un adjectif n'est pas le responsable de la forme du substantif ou de l'adjectif, de même il n'est pas responsable de la forme du proverbe qu'il emploie.

Ipsa facto, on comprend beaucoup mieux aussi la portée du critère de *comme on dit*. S'il convient à tous les proverbes, il ne leur est pourtant pas spécifique, comme le pense Gouvard (1996). Il satisfait aussi aux locutions figées non proverbiales :

Et, comme on dit, les carottes sont cuites.

Et, comme nous l'avions annoncé ci-dessus, il n'est nullement une manifestation directe du caractère collectif du *jugement* exprimé par le proverbe. Comme il ne s'applique pas aux dénominations simples :

?J'ai vu un «chien», comme on dit

et qu'il ne convient pas non plus aux séquences non figées :

?Et, comme on dit, il est huit heures

s'il s'applique aux proverbes et aux autres expressions figées, c'est parce qu'il s'agit de dénominations polylexicales.

3.2 Déproverbialisation

On peut à présent revenir au problème que constitue l'occurrence d'un proverbe avec des expressions les plaçant sous la responsabilité d'un particulier. Si l'on accepte notre caractérisation des proverbes comme étant des dénominations phrastiques, on peut expliquer quel élément se trouve modifié par une éventuelle insertion dans une structure du type *verbe d'opinion individuelle + proposition*. On sait qu'un proverbe en tant que phrase figée peut être, tout comme les autres expressions figées²³, défigé par un jeu formel (Grésillon et Maingueneau, 1984; Franken, 1995). Ces défigements portent sur l'aspect formel de la phrase. Il est aussi possible de ne pas modifier la forme, mais de toucher à l'autre aspect du proverbe, à son caractère de dénomination. Nous parlerons dans ce cas de *déproverbialisation*. La déproverbialisation est l'opération qui fait faire perdre au proverbe son côté dénominatif, pour ne lui laisser que son aspect de phrase. Le proverbe n'apparaît plus alors comme une phrase déjà construite, fixée, dont le sens est donné par avance, c'est-à-dire dont la prédication exprimée est acquise *a priori*, mais redevient une phrase comme les autres, qui présente à validation la combinaison SN-SV (s'il s'agit de cette structure). Un proverbe déproverbialisé est un proverbe qui perd son statut d'unité codée pour redevenir une simple phrase, une phrase générique, puisqu'ainsi que nous l'avons montré (1989a et 1994), les proverbes sont des phrases génériques. C'est ce qui se passe, nous semble-t-il, lorsqu'on place un proverbe sous la dépendance d'une expression de

²³ Pour une synthèse sur les locutions figées, voir Gross (1996) et Mejri (1996).

jugement individuel. C'est ainsi que le proverbe *L'argent ne fait pas le bonheur* s'il est inséré dans une structure du type *je trouve que / je crois que ...* en lecture non métalinguistique se trouve être déproverbialisé : il se voit privé de son statut de dénomination pour ne conserver que le statut d'une phrase générique, par laquelle, le locuteur-auteur donne son avis personnel sur le rapport entre le fait d'avoir de l'argent et celui d'être heureux. On peut ainsi maintenir qu'un proverbe n'est pas l'expression d'un jugement particulier, tout en rendant compte par le biais de la déproverbialisation des emplois du type *Je trouve que l'argent ne fait pas le bonheur*.

Un argument de poids peut être invoqué : lorsque la forme figée, qui, on le rappelle, est l'élément essentiel du caractère dénomiatif du proverbe, est trop éloigné, soit par la construction syntaxique, soit par un côté métaphorique, de la forme des phrases génériques, elle rend difficile l'insertion du proverbe dans une structure qui le subordonne à une opinion particulière. Pour une bonne et simple raison : une phrase dotée d'une telle forme se combine difficilement à une expression qui, elle, ne présente, au contraire, aucun caractère saillant. Le proverbe ne passe du coup pas l'étape de la déproverbialisation : l'aspect formel fait que le caractère dénomiatif n'est pas gommé dans l'histoire et donc interdit au proverbe de passer pour une phrase générique exprimée par un particulier. C'est pour cette raison que des combinaisons telles que :

?Je crois que qui trop embrasse mal étreint
?Je pense que qui trop embrasse mal étreint
?Je trouve que à petite cloche grand son
?Je crois que loin des yeux loin du coeur

paraissent mal formées²⁴, alors que des structures telles que :

Je trouve que l'argent ne fait pas le bonheur
Je crois qu'il vaut mieux prévenir que guérir
J'estime que prudence est mère de sûreté

ne suscitent aucun sentiment de rejet, parce que la forme du proverbe inséré se prête à une déproverbialisation sans heurts. Insistons bien sur un point : ce n'est pas le côté connu du proverbe qui est à l'origine de la difficulté d'apparaître comme jugement individuel déproverbialisé. *Prudence est mère de sûreté* est un proverbe bien connu, qui, accepte pourtant d'être présenté comme le jugement d'un particulier, parce que sa forme ne suscite pas de difficultés d'insertion.

Une difficulté reste à surmonter. Si notre explication est correcte, pourquoi certains proverbes acceptent-ils d'être déproverbialisés par tel ou tel verbe d'opinion et non par tel ou tel autre ? Pourquoi le proverbe *Prudence est mère de sûreté* peut-il apparaître avec *j'estime que* et non avec *je trouve que* :

J'estime que prudence est mère de sûreté
?Je trouve que prudence est mère de sûreté.

²⁴ Michaux (1996) avait entrevu ce facteur en notant qu'outre le caractère de prédication originelle, «le caractère métaphorique du proverbe influence lui aussi le type de lecture possible.»

La raison en est simple : c'est parce que la phrase générique obtenue ne peut être un complément de *je trouve que*, mais de *j'estime que ...* C'est à cette place que peuvent entrer en ligne de compte les analyses et les résultats obtenus par Michaux (1995 et 1996). Il est en effet naturel que chaque verbe d'opinion sélectionne son type de complément propositionnel et qu'un proverbe dégradé au niveau de simple phrase générique doive satisfaire aux conditions posées par le verbe enchâsseur.

CONCLUSION «INDIVIDUELLE»

Peut-on en tirer parti pour distinguer différentes classes de proverbes ? Nous laisserons pour aujourd'hui cette question ouverte : elle mérite d'être traitée avec le plus grand soin, notamment dans sa comparaison avec les différentes catégories de phrases génériques²⁵. Il nous suffit d'avoir atteint le but que nous nous sommes fixé : celui d'avoir montré, grâce à l'hypothèse de la dénomination et de l'opération de déproverbialisation, pourquoi les proverbes ne pouvaient pas être des jugements individuels, malgré des données apparemment contraires. Corollairement, se trouve confortée — et c'est réjouissant — la tendance «optimiste» qui estime que les proverbes forment une classe suffisamment homogène pour être définie linguistiquement.

RÉFÉRENCES

- ANSCOMBRE, J.C. (1989) : «Théorie de l'argumentation, topoï et structuration discursive», *Revue québécoise de linguistique*, 18, n° 1, pp. 13-56.
- ANSCOMBRE, J.C. (1990) : «Les syllogismes en langue naturelle ? Déduction logique ou inférence discursive ?», *Cahiers de linguistique française*, 11, pp. 215-240.
- ANSCOMBRE, J.C. (1994) : «Proverbes et formes proverbiales : valeur évidentielle et argumentative», *Langue française*, 102, pp. 95-107.
- ANSCOMBRE, J.C. (1995) : «La nature des topoï», in Anscombe J.C. (dir), *Théorie des topoï*, Paris, Kimé, pp. 49-84.
- ARNAUD, P.J.L. (1991) : «Réflexions sur le proverbe», *Cahiers de lexicologie*, 59-2, pp. 6-27.
- ARNAUD, P.J.L. (1992) : «La connaissance des proverbes français par les locuteurs natifs et leur sélection didactique», *Cahiers de lexicologie*, 60-1, pp. 195-238.
- ARNAUD, P.J.L. et R. MOON (1993) : «Fréquence et emploi des proverbes anglais et français», in Plantin C. (dir), *Lieux communs, topoï, stéréotypes, clichés*, Paris, Kimé, pp. 323-341.
- BERRENDONNER, A. (1981) : *Éléments de pragmatique linguistique*, Paris, Minuit.
- DUCROT, O. (1975) : «Je trouve que», *Semantikos*, 1,1, pp. 62-88.
- FOREST, R. (1996) : «Noms propres, idiotismes et polyphonie», *Bulletin de la Société de Linguistique de Paris*, t. XCI, 1, pp. 55-76.

²⁵ Nous espérons pouvoir mener une telle investigation sur les bases de notre approche des phrases génériques (Kleiber et Lazzaro, 1987; Kleiber, 1988 et 1989b).

- FRANKEN, N. (1995) : «Sur les détournements de proverbes», Projet ARC «Typologie textuelle et théorie de la signification», Rapport de recherches n° 3, Bruxelles, Université Libre de Bruxelles.
- GOUVARD, J.M. (1996), «Les formes proverbiales», *Langue française*, 110, pp. 49-63.
- GOUVARD, J.M. (à paraître), «Les proverbes dans *Le Paysan parvenu*».
- GROSS, G. (1996) : *Les expressions figées en français*, Paris, Ophrys.
- GRÉSILLON, A. et D. MAINGUENEAU (1984) : «Polyphonie, proverbes et détournement ou *Un proverbe peut en cacher un autre*», *Langages*, 73, pp. 112-125.
- KLEIBER, G. (1984) : «Dénomination et relations dénominatives», *Langages*, 76, pp. 77-94.
- KLEIBER, G. (1987) : *Du côté de la référence verbale. Les phrases habituelles*, Berne, Peter Lang.
- KLEIBER, G. (1988) : «Phrases génériques et raisonnement par défaut», *Le français moderne*, LXVI, 1/2, pp. 1-16.
- KLEIBER, G. (1989a) : «Sur la définition du proverbe», *Recherches Germaniques*, 2, pp. 233-252.
- KLEIBER, G. (1989b) : «Généricité et typicalité», *Le français moderne*, LXVII, 3/4, pp. 127-154.
- KLEIBER, G. (1994) : *Nominales*, Paris, A. Colin.
- KLEIBER, G. et H. LAZZARO (1987) : «Qu'est-ce qu'un SN générique ou Les carottes qui poussent ici sont plus grosses que les autres», in Kleiber G. (dir), *Rencontre(s) avec la généricité*, Paris, Klincksieck, pp. 73-111.
- MEJRI, S. (1996) : *Le figement lexical*, Thèse d'État, Tunis, Université de Tunis.
- MICHAUX, C. (1995) : «Parémies, collocations verbales et actes de parole ou comment les verbes aident à la classification et à la maîtrise des proverbes», Projet ARC «Typologie textuelle et théorie de la signification», Rapport de recherches n° 3, Bruxelles, Université Libre de Bruxelles.
- MICHAUX, C. (1996) : «Proverbe et jugement individuel : deux incompatibles ?», Projet ARC «Typologie textuelle et théorie de la signification», Rapport de recherches n° 4, Bruxelles, Université Libre de Bruxelles.
- NORRICK, N. (1985) : *How Proverbs mean. Semantic Studies in English Proverbs*, Berlin, Mouton.
- OLLIER, M.L. (1976) : «Proverbe et sentence : le discours d'autorité chez Chrétien de Troyes», *Revue des Sciences Humaines*, XLI, 163, pp. 329-357.
- RODEGEM, F. (1984) : «La parole proverbiale», in Suard F. et C. Buridant (dir), *Richesse du proverbe. Typologie et fonctions II*, Lille, Presses Universitaires de Lille, pp. 121-129.
- SCHULZE-BUSACKER, E. (1984) : «Proverbe ou sentence : essai de définition, in Di Sefano G. et R.G. Mc Gilliway (dir), *La locution*, Montréal, Éditions Ceres, pp. 134-167.
- SOMOLINOS RODRIGUEZ, A (1993) : «Arcaismos sintacticos en los proverbios franceses», *Paremia*, 1, pp. 55-63.
- SPERBER, D. et D. WILSON (1989) : *La pertinence*, Paris, Minuit.
- TAYLOR, A. (1931), *The Proverb*, Cambridge, Mass.

À LA RECHERCHE DE LA MÉMOIRE PERDUE. OU POUR UN DICTIONNAIRE HISTORIQUE DE L'ARABE

Taïeb BACCOUCHE

Université de Tunis I, Tunis, Tunisie

Quand on considère l'histoire de la lexicographie arabe, on est frappé par un paradoxe patent :

Nous sommes d'un côté en présence d'une histoire lexicographique et lexicologique fabuleuse qui surprend par la maturité et l'originalité du premier dictionnaire que la tradition attribue à El-Khalil (2^eS.H; 8^eS.J.)¹ ainsi que par la quantité, la qualité et la variété des lexiques et des dictionnaires conservés. Mais d'un autre côté, une tradition aussi riche n'a pas abouti jusqu'à nos jours à la confection ni d'un dictionnaire proprement étymologique, ni d'un dictionnaire historique. Cette lacune de taille représente à l'heure actuelle le point le plus faible de la lexicographie arabe au point de laisser à l'observateur l'impression que les lexicographes arabes, tant anciens que modernes se soucient moins de conserver la mémoire des mots de leur langue que de les consigner et de les expliquer tout simplement, répondant ainsi à un besoin essentiellement didactique. On pourrait donc légitimement se poser la question de savoir pourquoi cette carence. Y a-t-il eu des tentatives et pourquoi n'auraient-elles pas abouti ? Existe-t-il dans la tradition lexicographique arabe des indices d'une vision historique du vocabulaire ? En fait, les indices ne sont pas totalement absents. Il suffit d'examiner les termes utilisés dans le classement du vocabulaire arabe pour se rendre compte de l'existence d'une « conscience » historique du vocabulaire. Il est en effet de tradition depuis les premiers traités de linguistique arabe de distinguer dans le vocabulaire :

- الفصح [al-fasi:h] «le bon usage» conforme à la norme représentée notamment par le Coran, les dits du prophète Mohamed et la poésie classique (deux à quatre siècles après l'avènement de l'Islam selon qu'on est en zone citadine ou bédouine).

- الغريب [al-γari:b] «étrange», «hors usage», désignant en particulier des mots vieillis et sortis de l'usage ou même des expressions et des mots idiolectaux comme ceux attribués au prophète.

¹ *Al-ʿajn*, du nom de la première lettre d'après son classement. Ce dictionnaire étant classé selon l'articulation des phonèmes (laryngales ---> labiales). Toutes les combinaisons des phonèmes sont envisagées; ce qui permet à l'auteur de distinguer المستعمل [al- mustaʿmal] (unités faisant effectivement partie du système de la langue et المهمل [al-muhmal], (unités non utilisées par le système, mais représentant un potentiel latent et disponible).

- الدجيل [ad-daǧi:l] «intrus», désignant les emprunts qui sont souvent perçus comme des xénismes mais qui peuvent accéder au statut de *fasi:h* s'ils sont intégrés dans des schèmes arabes leur permettant de fonctionner normalement dans le système dérivatif arabe; dans ce cas, ils sont qualifiés de معرب [muʿarrab] «arabisés»; ils restent néanmoins *fasi:h* de seconde zone.

- المولد [al-muwallad] «généré» et المحدث [al-muhdaθ] «néologisme» désignent les mots créés après la fin de l'ère de la *fasa:ha* pour répondre aux besoins nouveaux d'une civilisation arabo-islamique en expansion.

On peut ajouter العامي «populaire», qui désigne soit des néologismes populaires, soit des mots arabes «corrompus» phonétiquement, morphologiquement ou sémantiquement.

Cette catégorie peut aller jusqu'au vulgaire السوقي [as-su:qij]. Toutes ces catégories qui se classent en marge du «bon usage» sont quelquefois réunies sous le terme عجمة [ʕuǧma] ou أعجمي [ʔaʕʕamij] «étranger, non arabe» ou «non conforme à la norme arabe».

D'autres termes sont utilisés comme مدموم؛ ردي؛ ضعيف؛ متروك etc., qui n'ajoutent aux diverses variantes que de simples nuances. Il y a là bien évidemment une conscience historique du vocabulaire englobant le vieillissement, la néologie et le contact des langues. Il est cependant paradoxal de constater que ce ne sont pas les linguistes arabes anciens qui ont formulé clairement cette conscience mais les penseurs, représentés brillamment sur ce point précis par l'historien sociologue Ibn Khaldoun (8^eS.H.; 14^eS.J.) qui considère dans ses fameux «Prolégomènes»² l'arabe citadin et l'arabe bédouin de son temps comme deux langues distinctes, différentes de l'arabe classique du temps du Prophète. Chacune de ces langues, fonctionne selon un système propre qu'il appelle ملكة [Malaka] et qu'on pourrait aujourd'hui traduire sans trop forcer par «compétence linguistique».

Pour répondre à la question : pourquoi les lexicographes arabes n'ont-ils pas poussé la réflexion à un niveau leur permettant de tirer profit de cette conscience historique et de jeter les bases d'une lexicographie historique ?, il est de tradition d'expliquer ce fait par le facteur religieux; le Coran étant un texte sacré, il est le modèle de la *fasa:ha* par excellence. Étant un texte atemporel, transhistorique, son support linguistique, l'arabe, est donc perçu comme en dehors du temps, échappant à l'historicité. Cette transcendance continue plus ou moins inconsciemment à hanter les lexicographes arabes jusqu'à nos jours, les empêchant de saisir la mémoire des mots dans sa dynamique.

Mais malgré cela, des tentatives plus ou moins timides ont été faites. Le rôle de pionnier dans ce domaine revient à un orientaliste allemand, Auguste Fischer (1865-1949), qui a le premier tenté d'élaborer un dictionnaire historique de l'arabe. L'idée a commencé à germer au milieu des années trente, à la suite de la création de l'Académie arabe du Caire, lorsque Fischer, membre de cette Académie, lui a présenté son projet; l'accord lui fut donné en 1938.

Ce travail avait l'ambition de couvrir cinq siècles (4^e → 9^eS.J.); 264 sources ont été à cette fin compulsées. Malheureusement, ce travail, qui a été partiellement relayé après la disparition de son promoteur, n'a pas abouti, faute de se voir transformé en une oeuvre de longue haleine, avec des structures et des moyens stables; cet échec traduit

² *al-muqaddima*, éd. Égypte, S.d. (ch.38,39,40,41), p. 554-562.

l'absence de conviction ou de volonté surtout politique à travers les structures théoriquement habilitées à entreprendre un travail de ce genre. Nous pensons en particulier aux Académies, à l'Alecso et à l'Union des universités arabes.

Pour illustrer cette carence, nous avons entrepris d'étudier le traitement réservé par les dictionnaires arabes à six mots du vocabulaire : quatre concernant les points cardinaux et deux, les deux dernières religions monothéistes, le Christianisme et l'Islam. Nous verrons pourquoi cette troisième paire n'est pas sans liens avec les deux autres qui forment un groupe de mots courants, actuels, très polysémiques et lourdement connotés tant en arabe qu'en français. Nous partirons des définitions données par le لسان العرب [lisan el-ʿarab](L), dictionnaire quasi-encyclopédique dû à Ibn Mandûr (+1311 J.C.), que nous comparerons à celles données par المنجد [el-munʿid] (M1=éd 1906 et M2=éd 1986) et المحيط [al-muhi:t] (Mht), le dernier né des dictionnaires arabes modernes³. Nous comparerons ces définitions à celles données par le *GLLF* (éd. 1971) et le *Robert* (éd. 1985) en vue de dégager la part réservée à la mémoire de ces mots dans les descriptions lexicographiques respectives.

1. شرق [ʃarq] «orient»

Mis à part le sens étymologique qu'évoque dans **L** la racine trilitère [ʃ rq] à savoir «fendre, éclater...», le sens principal attesté de *ʃarq*, nom verbal substantivé est «lieu où le soleil se lève» avec une variante morphologique sur le schème du nom de lieu مشرق [maʃriq] **M1** n'ajoute presque rien, tandis que **M2** en affine la formulation, l'étendant au point cardinal de l'est. **Mht** ajoute le sens plus moderne de «régions et pays de l'est». Pour ce dernier sens, le *GLLF* avance la date 1080 et les dates : 1949 pour *Proche-Orient*, début **XX^e** siècle pour *Moyen-Orient*, et 1968 pour *Extrême-Orient*.

	GLLF	Robert	Muht	M2	M1	L
I-1- Partie du ciel, côté de l'horizon où semble se lever le soleil.	+	+	+	+	+	+
2-Vx ou litt. = direction de l'orient, pt cardinal de l'est.	+	+	+	+	+	
3-Vx ou litt. = direction, orientation	+	+				
4- Fig. Féerie de lumière.	+					
5- Class. et litt. = commencement brillant, heureux.	+					
6- Reflet des perles.	+	+				
7- Fig. Couleur évoquant (6)	+	+				
II-1- Zone ou ensemble des pays de l'est, P-0, M-0, E-0.	+	+	+			
2- Habitants et nations de l'orient.	+	+				
III- Dans la franc-maçonnerie, loges de province.				+		

Tableau 1 : Orient - شرق [ʃarq]

³ Éd. Beyrouth, Paris, 1993, 3 vol.

Il est à remarquer que l'arabe ne distingue pas **orient** et **est**, qui n'ont pas forcément la même distribution en français alors que les deux variantes morphologiques arabes ont une distribution pratiquement identique.

En outre, les cases qui semblent vides en arabe, par comparaison aux sens français, sont en partie remplies par d'autres formes dérivées. Notons enfin que le sens politique né de la guerre froide n'est pas explicitée dans le dictionnaire français et est absent dans les dictionnaires arabes. Mais l'édition 92 du *Petit Robert* relève ce nouveau sens dans **est**.

Remarquons enfin que le dictionnaire arabe le plus récent, **Mht**, qui donne comme premier sens à شرق [ʃ arq], «soleil» (par métonymie), ne suit ni l'ordre logique ni l'ordre chronologique. Il ressort de cette comparaison que les dictionnaires français fournissent des descriptions beaucoup plus détaillées et datées.

2. غرب [ɣarb] «occident»

Le sens principal qu'évoque la racine trilitère [ɣrb] attesté dans tous les dictionnaires arabes est «limite», lié dans **L** aux sens «départ, éloignement, absence, disparition...». Le sens «limite» engendre «tranchant» s'agissant d'armes blanches. Celui d'«absence» engendre «obscurité». La forme غرب - ɣarb, nom verbal substantivé, avec sa variante morphologique de nom de lieu مغرب [maɣrib], signifie «lieu où le soleil se couche». **M2** précise cette définition en l'étendant au point cardinal de l'ouest. **Mht** ajoute le sens plus moderne de «pays situés à l'ouest» qu'il place en deuxième position au lieu de le placer en dernier lieu conformément à la chronologie d'évidence. Pour ce même sens, le *GLLF* date le sens «Partie ouest du continent européen», v.1119, puisqu'on parlait déjà d'*Empire d'occident* et d'*Église d'occident* par opposition à *Empire d'orient* et *Église d'orient*.

	GLLF	Robert	Muhit	M 2	M 1	L
1- Un des points cardinaux situé du côté de l'horizon où le soleil se couche.	+	+	+	+	+	
2- Partie ouest de l'Europe + ensemble des pays et des peuples qui l'occupent.	+	+				
3- Civil, culture des peuples de l'E.occidentale.	+					
4- Ensemble des États membres du pacte de l'Atlantique Nord.	+	+	+			
5- Dans une loge maçonnique, côté où se tiennent les surveillants	+					
6- Class. et fig. = déclin, ruine, catastrophe.	+					

Tableau 2 : Occident غرب [ɣarb]

Remarquons là encore que l'arabe ne distingue pas occident et ouest. On peut appliquer ici la même comparaison effectuée à propos d'*orient*. Cependant, le sens politique né de la guerre froide se retrouve dans les dictionnaires français : ensemble des États membres du pacte de l'Atlantique Nord, par opposition aux États de l'est de l'Europe et à ceux d'Asie, daté par le *GLLF*, 1959. Ce sens est ici explicite car *occident* peut se substituer à *Bloc de l'ouest* ou *pacte de l'Atlantique Nord* dans le discours politique moderne alors que *orient* ne se substitue pas à *Bloc de l'Est* ou *pacte de Varsovie*.

Ce sens, fréquent dans la littérature politique et médiatique arabe moderne, n'est pas attesté dans les dictionnaires arabes les plus récents.

Il ressort de ce qui précède que les dictionnaires arabes couvrent beaucoup moins de sens et ne respectent pas l'ordre chronologique.

3. شمال [ʃama:l] «nord»

Il est frappant de constater que **L** ne définit pas شمال mais l'évoque indirectement à partir d'un sens dérivé «vent du nord», que nous retrouvons en français daté par le *GLLF* début du XIII^e siècle.

L définit ce mot «vent soufflant du côté du pôle», قطب, employé seul, ne désigne que le pôle Nord. On y trouve quelques considérations astrologiques sur le point de départ de ce vent. On pourrait chercher la raison de ce paradoxe dans l'ambiguïté résultant de l'existence d'une variante en *i* [ʃimâl] désignant à la fois, le vent, le nord géographique et la gauche.

Le nord étant situé à la gauche de celui qui se place en face du soleil levant, ainsi [ʃa/imâl] s'oppose-t-il d'un côté à جنوب [ʃanu:b] «sud» et d'un autre côté à يمين [jami:n] «droite».

M1 et **M2** continuent à reproduire cette ambiguïté, alors que **Mht** consigne la spécialisation consacrée par l'usage : [ʃama:l] désignant le point cardinal et tout ce qui s'ensuit et [ʃima:l] la gauche géographique avec les mêmes connotations de mauvais augure que nous retrouvons dans *sinistre*.

Mht donne pratiquement tous les sens que nous trouvons en français mis à part le sens lié à la franc-maçonnerie. On y trouve également le sens récent de «pays développés et industrialisés» par opposition à جنوب [ʃanu:b] «pays pauvres» qu'on ne rencontre même pas dans l'édition 1992 du *Petit Robert*.

	GLLF	Robert	Muhit	M2	M1	L
1- Celui des 4 points card. qui correspond à la direction marquée par l'étoile Polaire	+	+	+			
+ le Nord = le vent du nord.	+	+	+	+	+	+
2- Lieu au Nord d'un point > (Nord)	+	+	+			
+ région du globe du côté nord		+				
3- Franc-maçonnerie = côté gauche						
en entrant en loge	+		+			
4- ≠Midi		+				
+ Pays industriels, développés.			+			

Tableau 3 : Nord شمال [ʃama:l]

4. جنوب [ʃanu:b] «sud»

L ne parle que du vent du sud, sec et chaud, avec quelques précisions astrologiques relatives à son point de départ. Il est paradoxal de remarquer que ce vent brûlant est considéré de bon augure parce qu'il souffle du يمين [Jami:n] «droite» (le sens de toutes les formes dérivées de la racine [jmn] évoque le bon augure, contrairement à celui de [ʃ ml] et sa variante étymologique [ʃ ?m]. M1 et M2 ajoutent à cela le sens géographique «point opposé au point nord».

Remarquons que Mht ne donne pas ici le sens «pays pauvres ou en développement» qu'il a cité par opposition dans [ʃ ama:l] «nord», probablement parce qu'il ne l'a pas introduit comme sens autonome mais comme exemple « حوار الشمال والجنوب » «dialogue Nord-Sud».

	GLLF	Robert	Muhit	M2	M1	L
1- Celui des 4 pts cardinaux corresp. à la direction de l'étoile polaire.	+	+	+	+	+	
2- Contrée située ds l'hémisphère Sud.	+	+				
3- Région située au sud dans un ensemble géographique.	+	+				
4- Ensemble des habitants d'une région sud.	+					
5- Adj.	+					
+ midi (de la France)		+				
+ vent du Sud		(+)	+	+	+	+

Tableau 4 : Sud جنوب [ʃanu:b]

Il ressort de l'opposition N-S que les tentatives d'actualisation existent comme le prouve **Mht** mais demeurent néanmoins non systématiques, puisque ce dictionnaire ne mentionne le nouveau sens «pays riches» et «pauvres» que dans l'article شمال [ʃama:l], et peu élaborées, puisqu'il ne lui accorde pas le statut de définition; se contentant du statut d'exemple.

5. إسلام [isla:m]

Nous retrouvons le sens étymologique de ce terme qui est un nom verbal substantivé «soumission, résignation à la volonté de Dieu» dans tous les dictionnaires arabes, ainsi que dans les dictionnaires français. Tous, excepté **L**, donnent bien évidemment le sens «religion des musulmans » et le sens dérivé «peuples qui pratiquent l'Islam et civilisation qui les caractérise». **Mht** évoque ce dernier sens à travers la définition de l'histoire musulmane; «histoire des peuples qui ont embrassé l'Islam». Quant à **L**, le sens «religion...» y est implicite à travers la définition non pas du concept mais du nom verbal, «le fait d'embrasser l'Islam».

	GLLF	Robert	Muhit	M2	M1	L
0- Soumission, Résignation à la volonté de Dieu.	+	+	+	+	+	+
1- Religion des Musulmans.	+	+	+	+	+	+
2- Peuples qui pratiquent l'islam						
+ civil . qui les caractérise	+	+	(+)	+	+	
+ militant islamiste → partisan de l'islamisme.						
		+				

Tableau 5 : Islam إسلام [ʔisla:m]

Il faut noter que le sens moderne d'Islam militant ou politique n'existe dans aucun dictionnaire arabe. En français ce sens se trouve de plus en plus sous la variante *Islamisme*. Mais seul le *Robert* relève ce nouveau sens sous la forme dérivée *islamiste* dans une citation du *Nouvel observateur* datée de 1984. Les dictionnaires arabes modernes semblent omettre ce sens en occultant l'opposition entre مسلم «musulman» et إسلامي «islamiste» pourtant si fréquente dans la production arabe moderne.

On a donc l'impression que cette carence dans les dictionnaires arabes est due à un blocage idéologique qui fait que seul ce qui est idéologiquement admis peut figurer dans les dictionnaires.

6. مسيحية [masi:hijja] «christianisme»

Le terme مسيحية dans le sens de «christianisme» n'existe ni dans **L** ni dans **M1** et **M2** pourtant élaborés par des pères. On le retrouve dans **Mht** comme synonyme de نصرانية, ce dernier terme qui veut dire littéralement «Nazaréen» se trouve dans le Coran sous la forme de نصارى [nasa:ra:] «chrétiens nazaréens» car les chrétiens les plus connus à l'époque

étaient surtout les nestoriens, les jacobites et les sabéens. Ce nom, comme le précise le *GLLF* est celui «donné aux premiers chrétiens d'après le nom de la ville de Nazareth en Palestine», et c'est encore ce terme qu'on retrouve dans les parlers arabes actuels. Ainsi, *مسيحية* est-il un terme plus récent utilisé fréquemment par les lettrés pour désigner le christianisme. Il est donc paradoxal de constater que les dictionnaires arabes, même les plus récents ne disent pas mot de cette évolution pourtant frappante vers une forme de spécialisation :

نصرانية : usage ancien d'origine religieuse qui se perpétue dans la tradition populaire sous la forme du pluriel *نصاري*

مسيحية : usage plus récent qui se généralise dans les écrits modernes.

	GLLF	Robert	Muhit	M2	M1	L
1- Religion (doctrine de J.C.).	+	+	نصرانية+	+	+	
2- Foi chrétienne de qn.	+		مسيحية+			

Tableau 6 : Christianisme *مسيحية* [masi:hijja] / *نصرانية* [nasra:nijja]

CONCLUSION

De l'examen de ces six termes et de leurs variantes, en arabe et en français, il est possible de dégager les principales remarques suivantes :

1- En français, les termes exprimant les points cardinaux s'organisent en paires qui s'opposent nettement : N-S; Ori-Occ (ou E-O).

Par contre, cette opposition nette géographiquement ne l'est pas linguistiquement en arabe. En effet, le même terme [ʃima:l] désigne dans les dictionnaires arabes le nord et la gauche qui est définie par rapport à l'orient et non par rapport au nord qui sert d'habitude de référence et de repaire.

Ces dictionnaires continuent à entretenir l'ambiguïté malgré la tendance très nette dans l'usage vers la spécialisation des variantes : [ʃama:l] «Nord» et [ʃima:l] «gauche».

2- Le sens politique né de ces oppositions dans l'usage moderne est généralement absent dans les dictionnaires arabes : Est-Ouest (guerre froide) et Nord-Sud (post-guerre froide).

Cependant l'opposition Nord-Sud, développée après la guerre froide, n'existe pas encore dans les dictionnaires français mais sera sûrement retenue dans les éditions à venir.

3- L'acquisition par le mot *Islam* et en particulier par sa variante *Islamisme* d'une connotation politique, qui l'oppose dans les écrits actuels non pas à *Christianisme* mais à *Occident* n'est pas mentionnée dans les dictionnaires arabes; elle l'est partiellement dans les dictionnaires français. Des livres et des colloques portent ce titre apparemment paradoxal «Islam et Occident». C'est la dimension politique récente de cette relation qui explique ou justifie la mise en opposition de ces deux entités.

4- La complexité des relations qui caractérisent les mots étudiés et leur caractère dynamique (nouvelle relation Islam-Occident) prouvent que les dictionnaires arabes ont

besoin d'évoluer dans leur approche afin de pallier leurs insuffisances et rattraper un retard patent.

Partant de ces constatations qui ont porté sur un échantillon très limité, nous avançons quelques propositions pour une méthodologie de la dimension historique du dictionnaire arabe :

1- Jusque là, la langue arabe, pour des raisons religieuses et idéologiques, a sombré dans une a-historicité bloquant toute initiative rendant compte de son évolution naturelle. Ce blocage était d'autant plus important que toutes les tentatives de descriptions historiques ont tourné court parce qu'elles ont voulu partir du Coran et de la littérature préislamique; d'où l'importance du point suivant : au lieu de suivre le cours du temps en faisant des découpages plutôt arbitraires, ne vaudrait-il pas mieux suivre le chemin inverse en partant des synchronies actuelles où l'on dispose de matériaux fiables, faciles à observer et à analyser, évitant ainsi les conjectures souvent liées à la méthode opposée et remonter petit à petit le temps vers les origines. C'est à partir de l'observable et du tangible qu'on pourrait façonner des outils méthodologiques fiables, capables de dépasser le retard accumulé et d'éclairer la mémoire obscurcie par tant de retard. Ainsi redonnerait-on à la langue une historicité qui lui revient de droit.

2- Il serait souhaitable que l'on procède par des descriptions portant sur des champs lexicaux où il serait plus aisé de dégager les valeurs respectives des items lexicaux et de reconstituer les structures sémantico- logiques organisant leur polysémie.

Les domaines de la recherche terminologique fondée sur l'évaluation du patrimoine lexical, la collecte des usages en cours dans les différentes régions et sur les possibilités offertes par le système, présentent aux lexicographes des champs bien structurés, facilitant ainsi cette recherche à rebours.

3- Si l'historicité est réhabilitée et les blocages idéologiques et religieux dépassés, les tentatives hésitantes d'actualisation observées dans certains dictionnaires récents pourrait être renforcées et les acquis méthodologiques de la lexicographie moderne bien mis à profit pour raviver et entretenir la mémoire des mots de l'arabe.

Remerciements

Je remercie mon ami Salah Mejri qui, après lecture et discussion de ce travail, a contribué à enrichir ses conclusions.

DÉNOTATION ET PROBLÈMES DE POLYSÉMIE DANS L'ÉLABORATION D'UN DICTIONNAIRE ÉLECTRONIQUE FRANÇAIS-ARABE

Bassam BARAKÉ

Université Libanaise, Tripoli, LIBAN

Au cours de mon travail sur un dictionnaire français-arabe¹, j'ai affronté maintes fois des problèmes de polysémie dans l'une et l'autre langue. Les cas de mots monosémiques ne sont pas très fréquents et, sur le plan de l'analyse contrastive, ils ne posent pas de problèmes majeurs, sémantiques ou syntaxiques². Mon propos ici se limite à la présentation des problèmes de dénotation et de polysémie dans une perspective de comparaison entre le français et l'arabe, cela à partir de l'analyse de quelques exemples, analyse dont le but est de présenter certains types de différences linguistiques et culturelles entre les deux langues et de délimiter les traits sémantiques, syntaxiques et sociolinguistiques qu'il faudrait prendre en compte dans l'élaboration d'un dictionnaire français-arabe.

* * *

Quelques remarques préliminaires

Le dictionnaire dont il s'agit est conçu dans l'optique du décodage de la langue source, le français, et de l'encodage en langue cible, l'arabe (dictionnaire de version français-arabe, pour usagers de langue arabe).

Nous adoptons l'arabe classique tel qu'il est actualisé de nos jours. Le canevas de départ est un dictionnaire du français contemporain et les équivalents arabes appartiennent généralement à l'arabe contemporain. Comme tout dictionnaire, mono- ou bilingue, est nécessairement dérivé d'une base de données, la tâche est relativement facile pour le français (les travaux dictionnaires, monolingues ou encyclopédiques, sont, dans notre cas et pour l'usage actuel de la langue, plus développés dans la langue source que dans la langue cible).

¹ Il s'agit du *Dictionnaire Larousse Français - Arabe* à paraître en décembre 1997.

² Les problèmes rencontrés avec les vocables monosémiques concernent moins les équivalences entre le français et l'arabe que les spécificités culturelles de la langue arabe (homonymie, synonymie, polylexie, etc.).

Si l'équivalence peut être parfaite lorsqu'il s'agit de termes techniques à caractère monosémique, certains cas de monosémie française trouvent une panoplie de «synonymes» arabes. Cela est dû à deux faits : d'une part, la communication entre les chercheurs et les écrivains arabes est difficile, pour des raisons d'ordre géographique et/ou politiques; d'autre part, la notion est généralement d'origine étrangère (une science ou un concept nouveau) et les locuteurs la désignent en arabe avec des termes qui reflètent la compréhension et l'origine «scientifique» de chacun (ex : le terme «linguistique» a connu quelque cinq équivalents arabes).

Vu les progrès réalisés par les moyens informatiques, principalement en ce qui concerne les supports, les outils, la mémoire et les possibilités de stockage et de traitement, nous pouvons envisager d'établir une base de données (sur CD-ROM ou sur un disque dur) dont le volume, les instruments d'utilisation et les voies de consultation dépassent de loin ce que nous pouvons faire sur le papier. Ainsi, les exemples que nous allons présenter sont traités dans la perspective de l'élaboration d'un dictionnaire électronique français-arabe.

Le dictionnaire bilingue, qu'il soit électronique ou sur papier, établit une relation entre les vocables de deux langues différentes. Il n'a pas à s'occuper du référentiel, dans le sens que sa destinée n'est pas d'être un dictionnaire encyclopédique; il confronte deux lexiques et deux découpages du monde, mais s'il s'intéresse d'abord et avant tout au découpage en signes, il ne pourrait pas négliger le côté référentiel, surtout lorsqu'il y a divergence culturelle entre les deux langues.

* * *

Prenons les deux vocables suivants :

fleuve et rivière

Le *Petit Robert* donne (entre autres) les définitions suivantes :

fleuve : 1. Grande rivière

rivière : I. 1. Cours d'eau naturelle de moyenne importance

Maupassant : «La rivière [l'Oued Saïda], fleuve là-bas, ruisseau pour nous, s'agite dans les pierres sous les grands arbustes épanouis».

Nous partons du principe fondamental qu'un dictionnaire bilingue français-arabe, sur papier ou électronique, est censé donner un synonyme en arabe de l'adresse en langue française. Mais, si nous prenons en considération le fait qu'une langue a pour rôle de décrire le réel, de le découper, nous aboutissons à la conclusion que les problèmes d'équivalence se posent sur deux plans : le plan de la réalité (niveau socio-culturel) et le plan de la langue. D'une culture à l'autre, il y a des différences de «vision» de la réalité et, par conséquent, des différences de vocables la désignant.

Dans les dictionnaires français-arabe, il existe un seul équivalent arabe du mot «fleuve», c'est «*nahr*».

Ce même mot arabe est aussi l'équivalent du mot «rivière». Celui-ci partage avec «ruisseau» un autre équivalent arabe : «*jadwal*».

fleuve	<i>nahr</i>
rivière	<i>nahr et jadwal</i>
ruisseau	<i>jadwal</i>

Ce qui nous importe ici c'est que là où le locuteur français dispose de deux mots (fleuve et rivière) pour désigner deux états d'un cours d'eau (de moyenne ou de grande importance), le locuteur arabe ne dispose que d'un seul mot. Les Égyptiens prennent conscience de l'incapacité du mot à désigner la majesté et l'immensité du Nil et l'appellent «mer du Nil».

Le recours aux «champs» syntaxique et sémantique rendra-t-il compte de cette différence (différence relative à une polylexie française) ? Ici, les traits syntactico-sémantiques sont indispensables pour l'analyse du lexique français (pour définir le signifié du vocable et déterminer son comportement syntaxique), mais ils ne sont pas suffisants pour décrire la polysémie de leur synonyme arabe. Dans un dictionnaire français-arabe (sur papier ou électronique), il faudrait rendre compte d'une telle différence entre le français et l'arabe et marquer l'équivalent de «fleuve» du trait «grand» et celui du mot «rivière» du trait «moyen». Cette différence n'étant pas due à l'emploi syntaxique dans l'une ou l'autre langue, le recours à la réalité socio-linguistique s'avère indispensable pour rendre compte des idiosyncrasies et des différences culturelles liés aux deux langues. «L'importance des écarts culturels nécessite, selon A. Rey, [...] une explicitation des différences, chaque fois qu'une simple équivalence lexicale ou idiomatique ne suffit pas» (Rey 1991 : 2865).

Prenons un exemple de polysémie du vocable français :

Campagne : I. Vaste étendue de pays découvert.
II. Étendue de terrain (zone de combat)
État de guerre, les combats
Travaux, action de communication limités à une période déterminée
(campagne de fouilles archéologiques, campagne de pêche, campagne électorale, campagne de publicité)

La polysémie du mot français est présentée par une polylexie arabe. Ici, plusieurs lexèmes arabes correspondent à plusieurs sens d'un seul lexème français. C'est donc la polylexie de la langue cible qui articule la polysémie de la langue source. Chacune de ces deux acceptions est traduite par un mot différent (*rîf* et *hamlah*). La différence de lexème équivalent suffit à rendre compte de la polysémie de la source lorsqu'il s'agit d'un dictionnaire de version (usager arabe). La détermination du champ syntaxique et de celui des classes d'objets (pour l'équivalent arabe) est indispensable pour l'usager français (thème).

À la suite des travaux de Gaston Gross, Brigitte Lépinette définit la description lexicographique par l'analyse des comportements syntaxiques. Dans le cadre du traitement automatisé du langage, elle préfère distinguer, dans un premier temps, les différents comportements syntaxiques du terme.

«Dans un second temps, en partant de l'hypothèse selon laquelle il y a souvent homogénéité entre les comportements syntaxiques et sémantiques (autrement dit, une construction et un environnement lexical donnés correspondent à un lexème et à un seul), il paraît possible d'organiser, en regard de ces unités en LS déterminées par leur comportement syntaxique, la présentation des équivalents traductifs de ces dernières en LC» (Lépinette 1997 : 60).

En effet, le recours à la notion de traits syntaxiques s'avère un moyen efficace pour lever les ambiguïtés liées aux phénomènes polysémiques. Prenons les vocables suivants : *chasseur* et *pêcheur*; Les lexèmes «chasseur» et «pêcheur» se traduisent tous les deux par un seul mot arabe «*sayyâd*», ainsi que les verbes chasser et pêcher (*sâda*). La polysémie du vocable arabe (par rapport à la langue source) ne rend pas compte de tous les sèmes dénotatifs de chacun des deux lexèmes français. Le dernier-né des dictionnaires français-arabe, celui de Abdelnour, réagit à ce fait comme s'il considérait que l'équivalent de chasseur est l'hyperonyme de celui de pêcheur et propose pour le premier «*sayyâd*» et pour le deuxième «*sayyâdu samak*», c'est-à-dire qu'il ajoute au premier le mot «poisson» pour rendre compte que le pêcheur est un «chasseur» de poissons. Cependant, en arabe, le verbe *sâda* (et le substantif de la même racine *sayyâd*) s'appliquent aussi bien au premier lexème qu'au second. Abdelnour a eu recours à la collocation syntaxique pour traduire la spécificité et la différence d'un des deux lexèmes français. Mais la solution qu'il propose, d'une part, ne permet pas de rendre compte du contenu lexical de «chasseur» (l'usager arabe peut comprendre le mot comme étant un synonyme de pêcheur) et, d'autre part, n'indique pas leur classe d'objets ni leur domaine. La faute n'incombe pas à l'auteur, c'est la propriété de tous les dictionnaires sur papier. En effet, les possibilités offertes par l'informatique permettent de doter chacun des lexèmes des traits indispensables à la reconnaissance de son contenu sémantique et de ses traits distinctifs.

Passons maintenant à un problème lié à la connotation de l'équivalent arabe. En effet, les écarts linguistiques, qui font l'objet du dictionnaire bilingue dans son ensemble et concernent principalement le plan sémantique et syntaxique, traduisent parfois des écarts culturels plus ou moins importants, qui concernent non seulement le plan référentiel (désignation), mais surtout le plan connotatif (culture).

Dans un poème d'Apollinaire nous lisons ceci :

Le mai le joli mai en barque sur le Rhin
Des dames regardaient du haut de la montagne
Vous êtes si jolies mais la barque s'éloigne
Qui donc a fait pleurer les saules riverains
Or des vergers fleuris se figeaient en arrière
...

Si nous voulons traduire ce poème en arabe la première difficulté que nous rencontrons est le vocable «mai». Le dénoté est le mois, cela est élémentaire. Mais l'équivalent arabe n'est pas aussi simple. Nous avons deux mots : *'Ayyâr* et *Mayo*. Or, si tous les deux mots rendent compte du sens lexical, aucun ne traduit la portée significative de la phrase. Le poète utilise un réseau de mots qui expriment le bonheur de vivre et la lumière éclatante qui accompagne l'arrivée du printemps. Notre vocable renvoie à «dames», «jolies», «fleuris», etc. Le «mai» ne saurait donc se traduire que par un mot qui rende compte de ces

significations connotant la joie et la lumière. Il s'agit de «*nawwâr*» dont la racine (*nûr*) exprime à la fois la fête, la lumière et le rayonnement. Dans l'élaboration d'un dictionnaire français-arabe général, et pour une meilleure utilisation du vocable, il faudrait tenir compte des faits suivants :

- La polylexie arabe
- Le niveau de langue du vocable
- les sens connotatifs liés au terme de la langue source et à son équivalent (quitte à prendre en considération, dans ce dernier cas, le facteur diachronique et la mémoire culturelle des mots.

Comme le souligne Alain Duval, «L'équivalence parfaite exige un même niveau de dénotation (c'est-à-dire la référence à un même élément de la réalité extérieure), et un même niveau de connotation, c'est-à-dire le même réseau d'associations culturelles liées au terme dans les deux langues» (Duval 1991 : 2818-2819).

CONCLUSIONS

L'analyse contrastive qui se veut le premier pas vers l'élaboration d'un dictionnaire bilingue doit prendre en considération le sens dans lequel celui-ci est constitué (langue maternelle → langue étrangère ou l'inverse), le public (l'utilisateur) auquel il est destiné et l'usage qu'il en fera. Dans cette perspective, il serait possible de prendre comme point de départ le profil sémantico-syntaxique fourni par le dictionnaire monolingue (des deux langues), en l'adaptant en fonction des résultats de l'étude différentielle des deux langues.

La différence entre un dictionnaire bilingue et un dictionnaire monolingue réside dans le fait que le bilingue est censé tenir compte des différences structurelles qui peuvent exister entre les deux langues, aussi bien sur le plan de la grammaire, de la dénotation et de la syntaxe que sur le plan du découpage du monde, de la culture et des structures sociales. Le travail d'élaboration doit passer par une étape d'analyse contrastive dont les résultats doivent être incorporés, d'une façon ou d'une autre, dans le dictionnaire en question.

L'analyse contrastive nous montre que l'équivalence n'est pas réciproque et que la symétrie entre les deux langues n'est possible que dans des cas rares.

L'élaboration d'un dictionnaire bilingue devrait passer par l'analyse contrastive des deux langues à partir de la comparaison de deux bases de données. La phrase constituera l'unité de base de l'analyse de chacune des deux langues, et l'introduction d'un champ sémantique qui puisse rendre compte des spécificités du signifié de l'équivalent (en comparaison avec la source) me paraît indispensable.

L'analyse contrastive de certains vocables de la paire de langues concernée révèle que la réalité dénotée par la langue source ne fait pas toujours partie de l'univers culturel des locuteurs de la langue cible, ou n'est pas reconnu en tant que tel par la majorité d'entre eux (comme nous l'avons vu avec l'exemple de «fleuve» et de «rivière»). Le réel dénoté n'existe que dans l'univers culturel et le lexique de la langue source et le vocable n'a de réalité que dans cette langue. C'est la conclusion à laquelle aboutit Alain Rey dans son analyse des contenus culturels dans les dictionnaires bilingues. Il soulève le problème du découpage référentiel d'univers culturellement différents.

«Les différences de découpage, écrit-il, proviennent a) de la nature des référents, b) des constructions conceptuelles effectuées par la culture à leur égard, c) de la prise en charge de ces conceptualisations par un usage d'une langue, lui même conditionné par les moyens — morphologiques, syntagmatiques, sémantiques... — qu'une langue met à la disposition d'une culture» (Rey 1991 : 2869).

Dans l'élaboration d'UN dictionnaire bilingue, l'analyse syntactico-sémantique fondée sur l'emploi est indispensable pour définir l'usage dans l'une et l'autre langue. Cependant, le recours aux champs conceptuel, culturel et socio-linguistique est lui aussi une nécessité.

RÉFÉRENCES

- CADIOT, Pierre et Benoît HABERT (1997) : «Aux sources de la polysémie nominale», *Langue française*, Larousse, mars, n° 113.
- DUVAL, Alain (1991) : «L'équivalence dans le dictionnaire bilingue», F. J. Hausmann et alii (dir), *Dictionnaires, Encyclopédie internationale de lexicographie*, Tome troisième, Berlin et New York, Walter de Gruyter, pp. 2817-2824.
- LADMIRAL, Jean-René et Henri MESCHONNIC (dir) (1981) : «La traduction», *Langue française*, Larousse, septembre, n° 51.
- LÉPINETTE, Brigitte (1997) : «Le rôle de la syntaxe dans la lexicographie bilingue», H. Béjoint et Ph. Thoiron (dir), *Les dictionnaires bilingues*, AUFELF-UREF/Duculot, pp. 53-69.
- REY, Alain (1991) : «Divergences culturelles et dictionnaire bilingue», F. J. Hausmann et alii (dir), *Dictionnaires, Encyclopédie internationale de lexicographie*, Tome troisième, Berlin et New York, Walter de Gruyter, pp. 2865-2870.

UN MODÈLE HYBRIDE POUR L'EXTRACTION DES CONNAISSANCES : LE NUMÉRIQUE ET LE LINGUISTIQUE

Ismail BISKRI^(1et3); Jean-Guy MEUNIER⁽¹⁾;
Christophe JOUIS^(2et3)

(1) Laboratoire de l'ANalyse Cognitive de l'Information, Université du Québec à Montréal, Canada; (2) IDIST/CREDO, Université Charles de Gaulle-Lille 3, France; (3) LALIC-CAMS, Université de la Sorbonne-Paris IV, France

1. INTRODUCTION

De nos jours, un nombre croissant d'institutions accumulent très rapidement des quantités de documents qui ne sont souvent classés ou catégorisés que très sommairement. Très vite, les tâches de dépistage, d'exploration et de récupération de l'information présente dans ces textes, c'est-à-dire des «connaissances», deviennent extrêmement ardues, sinon impossibles. Pour y faire face, il devient nécessaire d'explorer de nouvelles approches d'aide à la lecture et à l'analyse de texte assistées par ordinateur (LATAO).

Du point de vue méthodologique, la question de l'extraction des connaissances dans les textes rencontre des difficultés épistémologiques sérieuses. En raison de sa nature sémiotique et langagière, le traitement informatique traditionnel d'un texte est de nature linguistique. Un texte est vu comme une suite de phrases qu'on doit soumettre à des analyseurs linguistiques. Cette approche semble tout à fait naturelle, elle correspond théoriquement au processus naturel de lecture d'un humain. Cependant, cette approche s'avère problématique dès lors qu'il s'agit d'une grande masse de données textuelles.

Dans ce cadre, le traitement d'un texte par ordinateur en appelle à des dépôts de connaissances préconstruites acquises via des enquêtes cognitives (analyse de protocole) auprès des experts ou puisées dans le répertoire encyclopédique du savoir partagé. Ceux-ci sont alors utilisés comme gabarit dans le dépistage et la reconnaissance. De plus, les systèmes experts qui opèrent dans ce domaine doivent être dotés des mécanismes habituels (moteur d'inférence, maintien de cohérences, tests de plausibilité, etc.) leur permettant d'effectuer des déductions et des tests d'hypothèses avec un haut niveau de confiance et de réussite. Les connaissances comportent des représentations d'objets, de propriétés, de relations d'événements et de situations propres à l'objet à traiter, en l'occurrence le contenu informationnel du texte. En possession de ce savoir, ce système informatique de type expert pourrait alors réussir à «comprendre» le texte et donc en extraire les connaissances. De nombreuses recherches ont d'ailleurs montré la nécessité d'avoir les connaissances de multiples niveaux (syntaxiques, psycholinguistiques, lexicales, sémantiques,

encyclopédiques, etc.) (Regoczei et al., 1988; Shaw & Gaines, 1988; Jacobs & Zernik, 1988; Moulin et Rousseau, 1990; Zarri, 1990).

Du point de vue de la lecture et de l'analyse de texte assistées par ordinateur (LATAO), le problème de l'extraction des connaissances d'un corpus textuel se présente de manière totalement différente. Il est en effet délicat de donner a priori à l'ordinateur, les connaissances que le texte avait pour fonction de transmettre sauf peut-être, pour celles qui sont de nature générale, encyclopédique ou technique. Dans le cadre de LATAO, la connaissance se trouve dans le texte lui-même et doit en être extraite. Et les techniques qui ont donné des résultats intéressants en IA sur de petits textes bien maîtrisés (scénario de restaurant, etc.) s'avèrent vite problématiques lorsqu'elles sont appliquées à des domaines dont on ignore en partie ou en totalité la teneur. Un texte contient normalement de nombreux énoncés originaux qui n'ont pas encore été lus et dont le contenu tant lexical, sémantique, qu'encyclopédique est inconnu au préalable par le lecteur, et qu'il découvrira dans le parcours du texte lui-même.

Le deuxième problème est de nature plus technique. Même si on possédait des analyseurs linguistiques raffinés et robustes pouvant décrire un texte selon ses diverses catégories linguistiques (morphologiques, syntaxiques, sémantiques, discursives) il faudrait prévoir que ce traitement prenne un certain temps. Dans la meilleure des situations, la technologie actuelle ne permet guère d'analyser des phrases en deçà de quelque 10 à 20 secondes par phrase. On peut imaginer le temps requis pour traiter des milliers de pages. La situation de LATAO ne permet pas ce type de traitement. Il faut modifier l'approche. Des stratégies, peut-être plus grossières dans leurs approches premières, permettent ultimement des extractions fines de connaissances. C'est dans cette perspective que nous explorons les approches par classification numérique et plus particulièrement les classifieurs de type connexionniste. Il nous semble que, dans le traitement de grande masse d'informations, il faut y aller comme en archéologie. Un bon archéologue ne commence pas directement sa fouille par le plus fin et le plus précis de ses outils. Au contraire, il commence sa recherche par un parcours général de son territoire. Il utilise pour ce faire des outils généraux (sonar, résonance magnétique, géomatique, etc.). Ce n'est qu'une fois qu'il a cerné le lieu potentiel des vestiges archéologiques qu'il en appelle à des outils plus fins. La pelle, la cuillère, la brosse, etc. Et ce n'est qu'à la fin qu'il prendra son microscope électronique. En d'autres termes deux grandes étapes sont nécessaires, une première étape utilisant un outil que nous dirons «bulldozer» pour classifier d'une manière grossière les données textuelles et ainsi permettre à un utilisateur de sélectionner dans une deuxième étape les parties du texte sur lesquels il veut extraire des connaissances d'une manière plus fine et ce au moyen de méthodes linguistiques.

2. STRATÉGIES NUMÉRIQUES

La littérature technique relative au traitement de l'information textuelle a montré qu'il était possible d'explorer des outils d'extraction des connaissances dans des textes (*data mining*). Or, l'extraction de connaissances peut être vue sous plusieurs angles. Dans notre perspective, elle n'est pas une «compréhension» du texte, ni une paraphrase, ni un rappel d'information, mais un processus de traitement classificatoire qui identifie des segments de textes qui contiennent un «même» type d'information. Autrement dit, l'extraction des connaissances est définie comme résultant d'une opération de classification fondée sur l'un ou l'autre critère d'équivalence.

Pour les chercheurs dans le domaine de LATAO, cette problématique n'est pas nouvelle. Dans la recherche antérieure, plusieurs techniques et méthodes ont déjà été proposées pour tenter d'organiser le contenu d'un texte en des configurations interprétables. Ces méthodes, souvent moins fines certes que les approches linguistiques et conceptuelles, n'en permettent pas moins un premier parcours général et robuste du texte. Elles sont en mesure, par exemple, d'identifier dans un corpus des classes ou des groupes de lexèmes qui entretiennent entre eux des associations dites de cooccurrence et donc de détecter leurs réseaux sémantiques. Et les recherches actuelles commencent d'ailleurs à les privilégier de plus en plus (Church et al., 1989; Lebart et Salem, 1988; Salton, 1988, etc.). Parmi les modèles les plus couramment utilisés, on trouve habituellement l'analyse des cooccurrences, l'analyse corrélacionnelle, l'analyse en composante principale, l'analyse en groupe, l'analyse factorielle, l'analyse discriminante, etc. Malgré le succès qu'elles ont obtenu, on a dû constater que ces méthodes particulières posent deux problèmes importants. Premièrement, les modèles classiques ne peuvent traiter que des corpus stables. Toute modification du corpus exige une reprise de l'analyse numérique. Ceci devient un problème majeur dans des situations où le corpus est en constante modification (par exemple les dépôts de l'automobile électronique). Deuxièmement, les types de résultats qu'ils produisent ne sont pas sans problèmes théoriques. Ils posent des problèmes d'interprétation linguistique importants (Church et Hanks, 1990). Les associations des mots dans les classes ne sont pas toujours facilement interprétables. Pourtant, malgré leurs limites, ces approches ont été reconnues des plus utiles pour l'extraction des connaissances et plus particulièrement les connaissances terminologiques. D'une part, ces stratégies classificatoires permettent une immense économie de temps dans le parcours exploratoire d'un corpus, et à ce titre, elles sont incontournables lorsqu'on est confronté à de vastes corpus textuels. D'autre part, elles servent d'indices pour détecter rapidement certains liens sémantiques et textuels. Cependant, lorsqu'associées à des stratégies linguistiques plus fines et intégrées dans des systèmes hybrides (i.e., avec analyseurs linguistiques d'appoint), elles livrent une assistance précieuse pour des analyses globales. Elles permettent un premier déblaiement général du texte. Peuvent alors suivre des analyses plus fines.

Les recherches récentes permettent de penser qu'on peut améliorer ces techniques de classification de l'information. En effet, de nouveaux modèles classificateurs dits émergentistes commencent à être explorés pour ce type de tâche. Ils ont pour fondement théorique que le traitement «intelligent» de l'information est avant tout associatif et surtout adaptatif. Parmi ces modèles dits «de computation émergente» on distingue les modèles «génétiques», markoviens (Bouchaffra et Meunier, 1993) et surtout connexionnistes. Parmi ces derniers, on trouve une grande variété de modèles, entre autres, les modèles matriciels linéaires et non linéaires, les modèles thermodynamiques et les modèles basés tantôt sur la compétition, tantôt sur la rétropropagation, mais surtout sur des règles complexes d'activation et d'apprentissage (Kohonen, 1982). Les principaux avantages de ces modèles tiennent au fait que leur structure parallèle leur permet de satisfaire un ensemble de contraintes qui peuvent être faibles et même, dans certains cas, contradictoires et de généraliser leur comportement à des situations nouvelles (le filtrage), de détecter des régularités et ce, même en présence de bruit. Outre les propriétés de généralisation et de robustesse, la possibilité pour ces modèles de répondre par un état stable à un ensemble d'inputs variables repose sur une capacité interne de classification de l'information.

Cependant, tous ces modèles classifieurs émergentistes opèrent sur des données bien contrôlées et qui toutes doivent être présentes au début et tout au long du traitement. De plus, ils exigent souvent divers paramètres d'ajustement qui relèvent souvent d'une description statistique du domaine. Il s'ensuit que les résultats de classification obtenus sont valides pour autant qu'ils portent sur les données bien contrôlées où peu de modifications sont possibles. Si, après la période d'apprentissage, pour quelque raison que ce soit, les systèmes sont confrontés à des données qui n'étaient pas prévues dans les données de départ, ils auront tendance à les classer dans les prototypes déjà construits, donc à produire une sous-classification.

Or, le domaine dans lequel nous opérons, à savoir le texte, présente précisément ce type de problème. Chaque nouvelle page peut contenir des informations que le système peut ne jamais avoir rencontrées, et donc qu'il ne peut se permettre de classer dans ses prototypes antérieurement construits. Il faut donc, outre la dynamique de l'apprentissage, un système qui soit aussi plastique.

3. LA GRAMMAIRE CATÉGORIELLE COMBINATOIRE APPLICATIVE DANS LE CADRE DE LA GRAMMAIRE APPLICATIVE ET COGNITIVE

La Grammaire Applicative et Cognitive (Desclés, 1990) postule trois niveaux de description des langues :

a- le niveau phénotypique (ou le phénotype) où sont représentées les caractéristiques particulières des langues naturelles (par exemple, l'ordre des mots, les cas morphologiques, etc.). Les expressions linguistiques de ce niveau sont des unités linguistiques concaténées, la concaténation est notée par : $'u_1-u_2-...-u_n'$;

b- le niveau génotypique (ou le génotype) où sont exprimés les invariants grammaticaux et les structures sous-jacentes aux énoncés du niveau phénotypique. Le niveau génotypique est structuré comme un langage formel appelé «Langage génotype»; il est décrit par une grammaire appelée «Grammaire applicative»;

c- le niveau cognitif où sont représentées les significations des prédicats lexicaux par des schèmes sémantico-cognitifs.

Les trois niveaux font appel à des formalismes applicatifs typés où l'opération d'application d'un opérateur à un opérande est considérée comme primitive. Les niveaux deux et trois s'expriment dans le formalisme de la logique combinatoire typée de Curry et Feys (1958). Cette logique fait appel à des opérateurs abstraits — appelés «combineurs» — qui permettent de composer intrinsèquement des opérateurs plus élémentaires entre eux (Desclés, 1990). Les combineurs sont associés à des règles d'introduction et d'élimination. Ceux que nous utiliserons dans cet article¹ sont **B**, **C***, avec les règles d'élimination (β -réduction) suivantes (U_1 , U_2 , U_3 sont des expressions applicatives typées) :

¹ Le combineur **C*** est souvent noté **T**.

$$\begin{aligned} ((\mathbf{B} U_1 U_2) U_3) &> (U_1 (U_2 U_3)) \\ ((\mathbf{C}_* U_1) U_2) &> (U_2 U_1) \end{aligned}$$

Le modèle de la Grammaire Catégorielle Combinatoire Applicative (GCCA) relie explicitement les expressions phénotypiques à leurs représentations sous-jacentes dans le génotype². Le système consiste en :

- (i) une analyse syntaxique des expressions concaténées du phénotype par une Grammaire Catégorielle Combinatoire;
- (ii) une construction à partir du résultat de l'analyse syntaxique d'une interprétation sémantique fonctionnelle des expressions phénotypiques.

Les Grammaires Catégorielles assignent des catégories syntaxiques à chaque unité linguistique. Les catégories syntaxiques sont des types orientés engendrés à partir de types de base et de deux opérateurs constructifs '/' et '\'.

- (i) N (syntagme nominal) et S (phrase) sont des types de base.
- (ii) Si X et Y sont des types orientés alors X/Y et X\Y sont des types orientés³.

Une unité linguistique u de type orienté X sera désigné par '[X : u]'.

Les deux règles d'application (avant et arrière) sont notées :

$$\begin{array}{ccc} [X/Y : u_1] - [Y : u_2] & & [Y : u_1] - [X\Y : u_2] \\ \text{-----} \rightarrow & ; & \text{-----} \leftarrow \\ [X : (u_1 u_2)] & & [X : (u_2 u_1)] \end{array}$$

Les prémisses dans chaque règle sont des concaténations d'unités linguistiques à types orientés considérées comme étant des opérateurs ou des opérands, la conséquence de chaque règle est une expression applicative avec un type orienté.

La Grammaire Catégorielle Combinatoire (Steedman, 1989) généralise les Grammaires Catégorielles classiques en introduisant des opérations de changement de type et des opérations de composition des types fonctionnels. Dans la GCCA les règles de la Grammaire Catégorielle Combinatoire de Steedman introduisent les combinateurs **B**, **C*** dans la séquence syntagmatique. Cette introduction permet de passer d'une structure concaténée à une structure applicative. Les règles de la GCCA sont :

Règles de changement de type :

² Dans le phénotype, les expressions linguistiques sont concaténées selon les règles syntagmatiques propre à la langue. Dans le génotype, les expressions sont agencées selon l'ordre applicatif.

³ Nous choisissons ici la notation de Steedman (1989) : X/Y et X\Y sont des types orientés fonctionnels. Une unité linguistique 'u' avec le type X/Y (respectivement X\Y) est considérée comme un opérateur (ou une fonction) dont l'opérande de type Y est positionné à droite (respectivement à gauche) de l'opérateur.

$$\begin{array}{l} [X : u] \\ \hline [Y/(Y \setminus X) : (C_* u)] \end{array} \xrightarrow{\mathbf{T}} \quad ; \quad \begin{array}{l} [X : u] \\ \hline [Y \setminus (Y/X) : (C_* u)] \end{array} \xleftarrow{\mathbf{T}}$$

$$\begin{array}{l} [X : u] \\ \hline [Y/(Y/X) : (C_* u)] \end{array} \xrightarrow{\mathbf{T}x} \quad ; \quad \begin{array}{l} [X : u] \\ \hline [Y \setminus (Y \setminus X) : (C_* u)] \end{array} \xleftarrow{\mathbf{T}x}$$

Règles de composition fonctionnelle :

$$\begin{array}{l} [X/Y : u_1] - [Y/Z : u_2] \\ \hline [X/Z : (B u_1 u_2)] \end{array} \xrightarrow{\mathbf{B}} \quad ; \quad \begin{array}{l} [Y \setminus Z : u_1] - [X \setminus Y : u_2] \\ \hline [X \setminus Z : (B u_2 u_1)] \end{array} \xleftarrow{\mathbf{B}}$$

$$\begin{array}{l} [X/Y : u_1] - [Y \setminus Z : u_2] \\ \hline [X \setminus Z : (B u_1 u_2)] \end{array} \xrightarrow{\mathbf{B}x} \quad ; \quad \begin{array}{l} [Y/Z : u_1] - [X \setminus Y : u_2] \\ \hline [X/Z : (B u_2 u_1)] \end{array} \xleftarrow{\mathbf{B}x}$$

Les prémisses des règles sont des expressions concaténées typées; les résultats sont des expressions applicatives (typées) avec éventuellement introduction d'un combinateur. Le changement de type d'une unité u introduit le combinateur C_* ; la composition de deux unités concaténées introduit le combinateur B .

Pour l'exemple suivant *La liberté renforce la démocratie* nous avons l'analyse suivante :

- | | | |
|-----|--|-----------|
| 1. | [N/N : <i>la</i>] - [N : <i>liberté</i>] - [(S \setminus N)/N : <i>renforce</i>] - [N/N : <i>la</i>] - [N : <i>démocratie</i>] | |
| 2. | [N : (<i>la liberté</i>)] - [(S \setminus N)/N : <i>renforce</i>] - [N/N : <i>la</i>] - [N : <i>démocratie</i>] | (>) |
| 3. | [S/(S \setminus N) : (C* (<i>la liberté</i>)))] - [(S \setminus N)/N : <i>renforce</i>] - [N/N : <i>la</i>] - [N : <i>démocratie</i>] | (>T) |
| 4. | [S/N : (B (C* (<i>la liberté</i>)) <i>renforce</i>)] - [N/N : <i>la</i>] - [N : <i>démocratie</i>] | (>B) |
| 5. | [S/N : (B (B (C* (<i>la liberté</i>)) <i>renforce</i>) <i>la</i>)] - [N : <i>démocratie</i>] | (>B) |
| 6. | [S : ((B (B (C* (<i>la liberté</i>)) <i>renforce</i>) <i>la</i>) <i>démocratie</i>)] | (>) |
| 7. | [S : ((B (B (C* (<i>la liberté</i>)) <i>renforce</i>) <i>la</i>) <i>démocratie</i>)] | |
| 8. | [S : ((B (C* (<i>la liberté</i>)) <i>renforce</i>) (<i>la démocratie</i>))] | B |
| 9. | [S : ((C* (<i>la liberté</i>)) (<i>renforce</i> (<i>la démocratie</i>)))] | B |
| 10. | [S : ((<i>renforce</i> (<i>la démocratie</i>)) (<i>la liberté</i>))] | C* |
| 11. | [S : <i>renforce</i> (<i>la démocratie</i>) (<i>la liberté</i>)] | |

Ainsi pour cet exemple, à l'étape 1 des types catégoriels sont assignés aux unités linguistiques. À l'étape 2, la règle (>) est appliqué aux unités linguistiques *la* et *liberté*. À l'étape 3 une règle de changement de type (>T) est déclenchée pour construire un opérateur (C* (*la liberté*)) à partir de l'opérande (*la liberté*). Cet opérateur est composé avec l'opérateur *renforce* à l'étape 4 par une opération de composition (>B) de façon à former un opérateur complexe (B (C* (*la liberté*)) *renforce*). Deux autres opérations respectivement (>B) et (>) suivent.

À l'étape 7 commence la réduction des combinateurs. Cette série de réduction se fait dans le génotype.

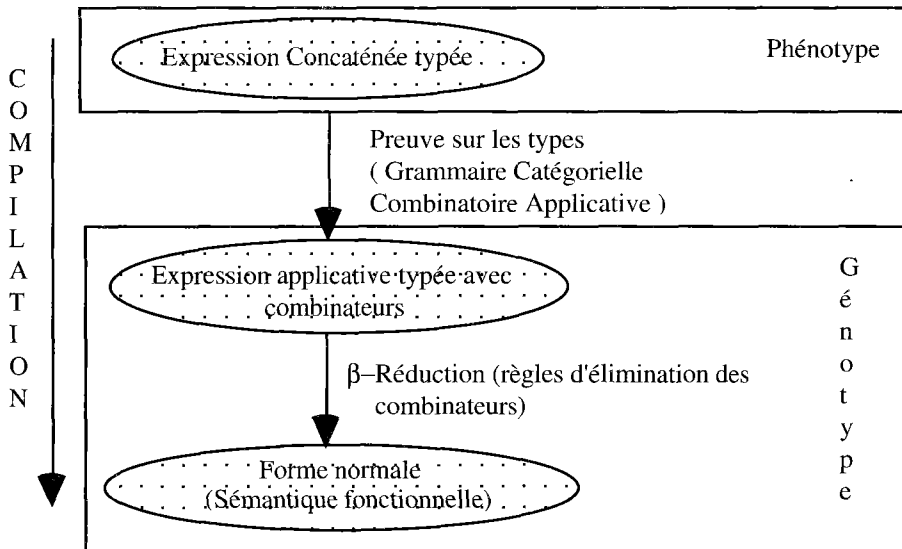
À l'étape 11 nous produisons la structure prédicative dont nous avons besoin pour aider le terminologue.

Un traitement complet basé sur la Grammaire Catégorielle Combinatoire Applicative s'effectue en deux grandes étapes :

(i) la première étape s'illustre par la vérification de la bonne connexion syntaxique et la construction de structures prédicatives avec des combinateurs introduits à certaines positions de la chaîne syntagmatique:

(ii) la deuxième étape consiste à utiliser les règles de β -réduction des combinateurs de façon à former une structure prédicative sous-jacente à l'expression phénotypique. L'expression obtenue est applicative et appartient au langage génotype. La GCCA engendre des processus qui associent une structure applicative à une expression concaténée du phénotype. Il nous reste à éliminer les combinateurs de l'expression obtenue de façon à construire la «forme normale» (au sens technique de la β -réduction) qui exprime l'interprétation sémantique fonctionnelle. Ce calcul s'effectue entièrement dans le génotype. Les formes applicatives obtenus en bout de parcours seront retenues pour être stockées dans des bases de données à des fins d'aide terminologique à l'utilisateur.

Le traitement fondé sur la GCCA prend la forme d'une compilation dont les étapes sont résumées dans la figure 1 :



4. LE MODÈLE HYBRIDE

Dans sa forme concrète le modèle hybride que nous proposons consiste en deux grandes étapes :

- Un filtrage numérique grossier du corpus qui permet de classifier et de structurer le corpus en des classes de termes qui serviront d'indices de régularités d'associations lexicales que le terminologue peut utiliser comme tremplin pour approfondir les étapes ultérieures d'interprétation, de construction de réseaux sémantiques, et finalement d'élaboration de ses fiches terminologiques. Une plate-forme réalisée au LANCI, en l'occurrence, la plate-forme ALADIN (Seffah et Meunier, 1995) permet d'exécuter une chaîne de traitement qui réalise un tel filtrage. La chaîne présente les étapes suivantes : elle commence par une gestion du document, suit alors une description morphologique (lemmatisation) et une transformation matricielle du corpus. Vient ensuite une extraction classificatoire par réseaux de neurones FUZZYART. Ainsi dans la première étape de sa gestion, le texte est reçu et traité par des modules d'analyse de la plate-forme ALADIN-TEXTE. Cette plate-forme est un atelier qui utilise des modules spécialisés dans l'analyse d'un texte. Dans un premier temps, un filtrage sur le lexique du texte est fait. Par divers critères de discrimination, on élimine du texte certains mots accessoires (mots fonctionnels ou statistiquement insignifiants, etc.) ou ceux qui ne sont pas porteurs de sens d'un point de vue strictement sémantique, et dont la présence pourrait nuire au processus de catégorisation, soit parce qu'ils alourdiraient indûment la représentation matricielle, soit parce que leur présence nuirait au processus interprétatif qui suit la tâche de catégorisation. Vient ensuite une description morphologique minimale de type lemmatisation.

Puis une transformation est opérée pour obtenir une représentation matricielle du texte. Cette transformation est encore effectuée par des modules d'ALADIN explicitement dédiés à cette fin. On produit ainsi un fichier indiquant pour tout lemme choisi sa fréquence dans chaque segment du texte. Suit ensuite un post-traitement pour construire une matrice dans un format acceptable par le réseau de neurone FUZZYART⁴.

Le réseau neuronal génère une matrice de résultats qui représentent la classification trouvée. Chaque ligne (ou vecteur) de cette matrice est constituée d'éléments binaires ordonnés. La ligne indique pour chaque terme du lexique original s'il fait ou non partie du prototype de la classe. Ainsi est créé un «prototype» pour chacune des classes identifiées. On dira alors que la classe numéro X est «caractérisée» par la présence d'un certain nombre de termes. Autrement dit, chaque classe identifie quels sont les termes qui se retrouvent dans les segments de textes qui présentent, selon le réseau de neurones une certaine similarité. Ainsi, les classes créées sont caractérisées, arbitrairement, par les termes qui sont présents également dans tous les segments du texte qui ont été «classifiés» dans une même classe.

Les résultats du réseau de neurones se présentent donc (avant interprétation) sous la forme d'une séquence de classes que l'on dira «caractérisées» par des termes donnés et incluant un certain nombre de segments.

⁴ Le réseau de neurones FUZZYART utilisé pour l'expérimentation d'ALADIN a été développé sur une plate-forme de programmation matricielle disponible sur le grand marché appelé MATLAB.

- Un traitement linguistique plus fin des segments sélectionnés selon les thèmes choisis par le terminologue. Ce dernier sélectionne donc des segments dont il veut une analyse plus fine et en extrait une représentation des connaissances plus structurée. Le terminologue peut décider pour un segment donné de focaliser son attention sur un terme donné et en construire son réseau sémantique. La Grammaire Catégorielle Combinatoire Applicative peut organiser les phrases dans lesquelles apparaît le terme, choisi par le terminologue, sous forme de structures prédicatives `Prédictat argument1 argument2... argumentn`. Ainsi pour une sélection de phrases on peut engendrer une liste d'expressions prédicatives. Nous pouvons avoir dans cette liste des structures prédicatives ayant des arguments en commun.

Par exemple, le cas suivant :

Prédictat1 argument1 argument2
Prédictat2 argument3 argument1

et là un terminologue comprendrait la relation sémantique entre les arguments 2 et 3 par rapport à l'argument 1.

Nous pouvons conserver une liste d'expression de la forme 'Prédictat argument1 argument2... argumentn' dans une base de données.

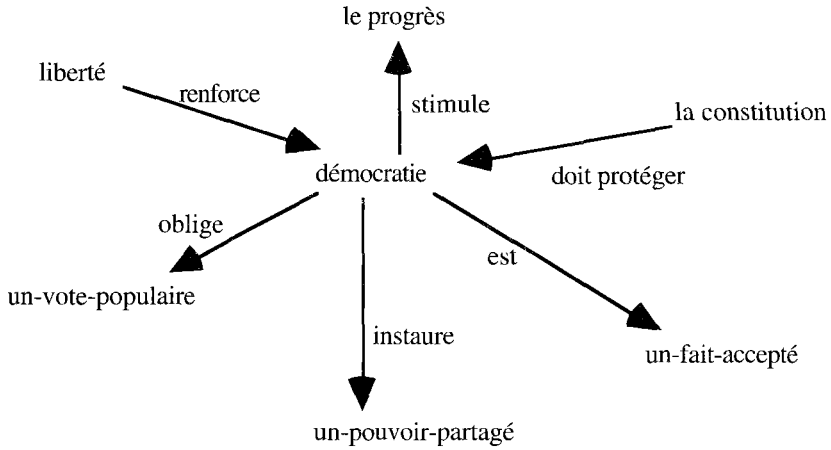
Le terminologue la consultant peut déduire le sens sémantique de chaque argument (donc terme) en ce sens qu'il peut en construire le réseau sémantique. Prenons les phrases suivantes :

La liberté renforce la démocratie.
La démocratie stimule le progrès.
La démocratie oblige un vote populaire.
La démocratie est un fait accepté.
La constitution doit protéger la démocratie.
La démocratie instaure un pouvoir partagé.

Les structures prédicatives de ces phrases obtenues par une analyse linguistique catégorielle sont respectivement :

renforce (la démocratie) (la liberté)
stimule (le progrès) (la démocratie)
oblige (un-vote-populaire) (la démocratie)
est (un-fait-accepté) (la démocratie)
(doit protéger) (la démocratie) (la constitution)
instaure (un-pouvoir-partagé) (la démocratie)

un terminologue ayant dans sa base de données ces structures prédicatives pourra représenter le réseau sémantique de (la démocratie).



Remarque

De tels exemples cependant ne mettent pas en évidence la pertinence de l'approche par classifieurs numériques car il n'y a pas d'ambiguïté dans les exemples. Le mot pôle *démocratie* a un sens unique.

Prenons le mot pôle *ferme*

Nous pouvons avoir trois groupes possibles :

Groupe 1

Une démocratie ferme est l'espoir d'un peuple.
Un vote ferme est passé à l'assemblée.
C'est une position ferme du gouvernement.

Groupe 2

Les paysans ont occupé les fermes et les villages.
Toutes les fermes ont abandonné leur récolte.
Les fermes sont laissées à l'abandon.

Groupe 3

Le vote ferme la discussion.
La décision du gouvernement ferme les options.
Le président de l'assemblée ferme le vote.

C'est le traitement classifieur qui a permis de séparer les champs lexicaux en trois groupes. Mais c'est au moyen de la grammaire catégorielle combinatoire applicative qu'une analyse en profondeur sera faite.

5. CONCLUSION

Nous venons de présenter un modèle d'extraction des connaissances pour terminologues. L'idée d'associer des modèles linguistiques à des modèles numériques est très prometteuse. Elle est également très pertinente, en ce sens qu'elle associe la finesse d'analyse des méthodes linguistiques à la capacité des méthodes numériques d'absorber de gros corpus. L'ordre stratégique d'appliquer une méthode numérique avant de faire intervenir une méthode linguistique résulte du compromis nécessaire pour faire `cohabiter` ces deux approches. En effet, la méthode numérique est plus à même de `débroussailler` un gros texte et de permettre à un terminologue de soumettre des segments choisis à l'analyseur linguistique plus fin.

Ce type d'approche permet de solutionner une des critiques importantes qu'on fait classiquement aux approches de cooccurrence et collocation : leurs difficultés d'interprétation. Notre approche permet d'entrevoir des outils de raffinement de l'analyse des résultats livrés par les approches numériques trop grossières et générales. Il y a compensation entre les deux. Les Grammaires Catégorielles sont trop fines et donc trop lentes sur un corpus ample. Mais bien placées elles ne travaillent que sur des sous-corpus qui ont effectué un premier travail de désambiguïsation.

En fait cette approche oblige à effectuer la désambiguïsation dans un traitement différent de celui de la grammaire. La désambiguïsation joue sur la différentialité des contextes de l'ensemble d'un corpus (relation paradigmatique) alors que l'analyse catégorielle opère sur la dépendance des contextes immédiats (relation syntagmatique). Ainsi, le système n'est pas obligé de faire les deux analyses en même temps ou dans une même passe ce qui, pensons nous, le rend plus efficace.

Enfin la configuration des résultats telle que permise par les expressions prédicatives permet au lecteur analyste d'avoir une organisation plus limpide sur le plan ergonomique et cognitif. On envisage la possibilité de livrer ces résultats dans des réseaux structurés mais aussi dans des repertoires structurés genre dictionnaires, thesaurus, bases de connaissances, etc.

Enfin, signalons que nous travaillons également sur un autre modèle hybride numérique/linguistique, associant réseaux de neurones et le modèle d'exploration contextuelle (système SEEK) (Jouis, 1993; Jouis, Biskri, Desclés, Meunier et al., 1997). Ce système permet de détecter des liens sémantiques de types cinématiques ou dynamiques (mouvements d'objets, changements d'états d'objets, relations de causalité, recherche des contextes définitoires d'un terme, etc.). Cela permettra d'élargir les interprétations par le terminologue des niveaux descriptifs statiques du domaine vers des descriptions évolutives.

RÉFÉRENCES

- BISKRI, I. (1995) : *La Grammaire Catégorielle Combinatoire Applicative dans le cadre de la Grammaire Applicative et Cognitive*, Thèse de doctorat, EHESS, Paris.
- BISKRI, I. et J.P. DESCLÉS, (1995) : «Applicative and Combinatory Categorical Grammar (from syntax to functional semantics)», *Colloque RANLP*, Bulgarie 1995.
- BOUCHAFFRA, D. & G. MEUNIER (1993) : «Theory and Algorithms for Analysing the Consistent Region in Probabilistic Logic», *An International Journal of Computers & Mathematics with Applications*, vol. 25, n° 3, February, edit. Ervin Y. Rodin, Published by Pergamon Press.
- BURR, D. J. (1987) : «Experiments with a Connectionist Text Reader», *IEEE First International Conference on Neural Networks*, San Diego, pp. 717-724.
- CARPENTER, G. & G. GROSSBERG (1991) : «An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition», *Neural Networks*, 4, pp. 493-504.
- CHEESEMAN, P., SELF, M., KELLY, J., STUTZ, J., TAYLOR, W. & D. FREEMAN (1988) : «Bayesian Classification», *Proceedings of AAAI 88*, Minneapolis, pp. 607-611.
- CHURCH, K., GALE, W., HANKS, P. & D. HINDLE (1989) : «Word Associations and Typical Predicate-argument Relations», *International Workshop on Parsing Technologies*, Carnegie Mellon University, Aug. 28-31.
- CHURCH, K. W. & P. HANKS (1990) : «Word Association Norms, Mutual Information, and Lexicography», *Computational Linguistics*, 16, pp. 22-29.
- CURRY, B. H. & R. FEYS (1958) : *Combinatory logic*, Vol. I, North-Holland.
- DELANY P. & P. LANDOW (Eds) (1993) : *The Digital Word : Text Based Computing in the Humanities*, Cambridge, Mass, MIT Press.
- DELISLE, S. (1994) : *Text Processing Without A Priori Domain Knowledge : Semi Automatic Linguistic Analysis for Incremental Knowledge Acquisition*, PhD Thesis, Ottawa University.
- DESCLÉS, J. P. (1990) : *Langages applicatifs, langues naturelles et cognition*, Paris, Hermès.
- DESCLÉS, J. P. & I. BISKRI (1996) : «Logique combinatoire et linguistique : Grammaire Catégorielle Combinatoire Applicative», *Revue Mathématiques, Informatiques et Sciences Humaine*, Paris.
- FREY, S., REYLE, U. & C. ROHRER (1983) : «Automatic Construction of a Knowledge Base by Analysing Texts in Natural Language», *Proc. of IJCAI 83*, pp. 727-729.
- GARNHAM, A. (1981) : «Mental Models and Representation of Texts», *Memory and Cognition*, 9, pp. 560-565.
- GREFENSTETTE, G. (1992) : «Sextant : Exploring Unexplored Contexts for Semantic Extraction from Syntactic Analysis», *Proc. of the 30th Annual Meeting of the ACL*, pp. 324-326.
- GREFENSTETTE, G. (1992) : «Use of Syntactic Context to Produce Term Association Lists for Text Retrieval», *Proc. of SIGIR 92 ACM*, Copenhagen, June 21-24.

- GROSSBERG, S. & S. CARPENTER (1987) : «Self Organization of Stable Category Recognition Codes for Analog Input Patterns», *Applied Optics*, 26, pp. 4919-4930.
- JACOBS, P. & U. ZERNIK (1988) : «Acquiring Lexical Knowledge from Text A Case Study», *Proceedings of AAAI 88*, St Paul, Min.
- JANSEN, S., OLESEN, J., PREBENSEN, H. & T. THARNE (1992) : *Computational approaches to text Understanding*, Copenhagen, Museum Tusulanum Press.
- JOUIS, C. (1993) : *Contributions à la conceptualisation et à la modélisation des connaissances à partir d'une analyse linguistique de textes. Réalisation d'un prototype : le système SEEK*, Thèse de doctorat, EHESS, Paris.
- JOUIS, C., BISKRI, I., DESCLÉS, J. P., LE PRIOL, F., MEUNIER, J.-G., MUSTAFA, W. & G. NAULT (1997) : «Vers l'intégration d'une approche sémantique linguistique et d'une approche numérique pour un outil d'aide à la construction de bases terminologiques», *Actes du colloque JST97*, Avignon, France.
- KOHONEN, T. (1982) : «Clustering, taxonomy and topological Maps of Patterns», *IEEE Sixth International Conference on Pattern Recognition*, pp. 114-122.
- LEBART, L. & A. SALEM (1988) : *Analyse statistique des données textuelles*, Paris, Dunod.
- LIN, X., SOERGEL, D. & G. MARCHIONINI (1991) : «A Self Organizing Semantic Map for Information Retrieval», *SIGIR 91*, Chicago, Illinois.
- MEUNIER, J.-G. (1996) : «Théorie cognitive : son impact sur le traitement de l'information textuelle», V. Riale et D. Fisette, *Penser l'esprit des sciences de la cognition à une philosophie cognitive*, Presses de Université de Grenoble, pp. 289-305.
- MOULIN, B. & D. ROUSSEAU (1990) : «Un outil pour l'acquisition des connaissances à partir de textes prescriptifs», *ICO*, Québec 3 (2), pp. 108-120.
- REGOCZEI, S. et al. (1988) : «Creating the Domain of Discourse : Ontology and Inventory». Gaines & Boose (Eds), *Knowledge Acquisition Tools for Experts and Novices*, Academic Press.
- REGOCZEI, S. & G. HIRST (1989) : On extracting knowledge from Text. Modeling the Architecture of Language Users. (TR CSRI 225), Computer Systems Research Institute, University of Toronto.
- SALTON, G. (1988) : «On the Use of Spreading Activation», *Communications of the ACM*, vol 31 (2).
- SALTON, G., ALLAN, J. & C. BUCKLEY (1994) : «Automatic Structuring and Retrieval of Large Text File», *Communications of the ACM*, 37 (2), pp. 97-107.
- SEFFAH, A. & J.-G. MEUNIER (1995) : «ALADIN : un atelier orienté objet pour l'analyse et la lecture de textes assistée par ordinateur», *International Conference on Statistics and Texts*, Rome.
- SHAUMYAN, S. K. (1987) : *A Semiotic Theory of Natural Language*, Bloomington, Indiana Univ. Press.

- SHAW, M. L. G. & B. R. GAINES (1988) : «Knowledge Initiation and Transfer Tools for Expert and Novices». Boose & Gaines (Eds), *Knowledge Acquisition Tools for Expert Systems*, Academic Press.
- STEEDMAN, M. (1989) : *Work in progress : Combinators and grammars in natural language understanding*, Summer Institute of Linguistic, Tucson University.
- TAPIERO, I. (1993) : *Traitement cognitif du texte narratif et expositif et connexionnisme : expérimentations et simulations*, Université de Paris VIII.
- THRANE, T. (1992) : «Dynamic Text Comprehension», J. O. S. Jansen, H. Prebensen, T. Thrane (Eds), Copenhaguen, Museum Tuscalanum Press.
- VERONNIS, J., IDE, N. M. & S. HARIE (1990) : «Utilisation de grands réseaux de neurones comme modèles de représentations sémantiques». *Neuronimes*.
- VIRBEL, J. (1993) : «Reading and Managing Texts on the Bibliothèque de France Stations», P. Delany & P. Landow (Eds), *The Digital Word : Text Based Computing in the Humanities*. Cambridge, Mass., MIT Press.
- YOUNG, T. & T. CALVERT (1987) : *Classification, Estimation, and Pattern Recognition*, Amsterdam, Elsvier.
- ZARRI, G. P. (1990). «Représentation des connaissances pour effectuer des traitements inférentiels complexes sur des documents en langage naturel», Office de la langue française (Ed.), *Les industries de la langue. Perspectives 1990*, Gouvernement du Québec.

DÉVELOPPEMENT DE LEXIQUES À GRANDE ÉCHELLE

Pierrette BOUILLON, Sabine LEHMANN, Sandra MANZI et
Dominique PETITPIERRE

ISSCO, Université de Genève, Genève, Suisse

1. INTRODUCTION

Cet article donne un aperçu général de trois descriptions linguistiques développées en parallèle pour l'anglais, le français, l'allemand et l'italien, à l'aide de l'outil morphologique *mmorph*. Il insiste surtout sur sa convivialité et ses diverses possibilités.¹

mmorph a été conçu dans le cadre du projet européen *Multext* («Multilingual Text Tools and Corpora», Projet LRE 62-050, 1994-1996). Ce projet s'est donné pour objectif de développer un ensemble d'outils pour le traitement de textes multilingues, destinés à des études de corpus ou à des applications dans le domaine du TALN. À son terme, il offre ainsi aux chercheurs des *corpus* et des *outils* pour les *annoter*. Ces derniers comprennent un *segmenteur* (qui segmente le texte en ses différentes unités : paragraphes, phrases et mots), un *analyseur morphologique* (*mmorph*) et diverses *descriptions linguistiques* (qui permettent de faire l'analyse morphologique du mot), un *étiquetteur probabilistique* (ou *tagueur*) (qui prend en entrée des mots avec des étiquettes ambiguës et lui attribue la plus probable en fonction de son contexte) et, enfin, un *aligneur* (qui établit des correspondances entre les diverses unités, paragraphes, phrases et mots, de deux textes de deux langues différentes). L'outil d'analyse morphologique que nous décrivons ici ne constitue donc qu'un des maillons dans la chaîne de traitement des corpus.

Dans la suite, nous décrivons d'abord l'outil *mmorph*, puis illustrons ses possibilités à l'aide d'exemples concrets tirés des descriptions linguistiques des différentes langues. Nous terminons en donnant une vue d'ensemble des lexiques.

2. L'OUTIL *mmorph*

mmorph est un outil qui permet de décrire la morphologie des mots d'une langue. Cette description est interprétée pour produire une base de données de mots fléchis. Celle-ci peut ensuite être manipulée de différentes façons, par exemple pour associer un mot fléchi à des informations morphologiques (analyse) ou pour dériver des formes fléchies à partir de la forme de base d'un mot et d'indications morphologiques (Petitpierre et Russell, 1995).

¹ Nous remercions Susan Armstrong et Graham Russell qui ont participé à ce travail et Suissetra qui a contribué au financement du projet.

mmorph est basé sur le paradigme de la morphologie à deux niveaux (Koskenniemi, 1984; Karttunen et Wittenburg, 1983 et Ruessink, 1989), avec l'addition de traits et de règles qui décrivent la syntaxe interne des mots (Pulman et Hepple, 1993; Ritchie et al., 1992).

Pour *mmorph*, la **description morphologique** d'une langue comporte quatre parties, a) un ensemble de déclarations, b) un ensemble d'entrées lexicales, c) un ensemble de règles de production lexicale, et d) un ensemble de règles morphographémiques (orthographiques).

a. Les **déclarations** définissent et nomment certains concepts utilisés dans la description morphologique (types, traits, valeurs, alphabets, etc.), ce qui permet au programme de vérifier la cohérence de la description morphologique, et à l'utilisateur de travailler avec une notation adaptée au cadre linguistique dans lequel il opère.

b. Les **entrées lexicales (lexique)** associent une description morphosyntaxique à des mots complets ou partiels (racines ou affixes). La description morphosyntaxique est composée d'un type et d'un ensemble de traits (des paires *attribut=valeur*). Les mots sont spécifiés par une chaîne de symboles lexicaux.

Le mot *étudiant*, par exemple, reçoit la description donnée en (1).

(1) Nom [genre=masculin nombre=singulier forme=surface type=2]
"étudiant" = "étudiant"

Il a le type Nom. Les traits *genre=masculin* et *nombre=singulier* indiquent qu'il est au masculin singulier; le trait *type=2* qu'il accepte une flexion du féminin (pour former *étudiante*) et le trait *forme=surface* qu'il constitue un mot complet sans adjonction d'affixe (contrairement à la racine «appel» du verbe *appeler*).

Le lexique décrit donc les mots complets ou partiels. Les règles de production définissent comment ils se combinent.

c. Les **règles de production lexicale** sont des règles de réécriture qui définissent ce qu'est un mot bien formé et comment les traits des morphèmes se combinent dans le mot. La règle (2), par exemple, crée un nom *pluriel* en concaténant un nom *singulier* (de type Nom) avec un suffixe nominal pluriel (de type Nom_suffixe), par exemple «mère + s» → *mères*. Elle précise que le nom pluriel aura le même *genre* que son homologue au *singulier* (le \$G indique une variable) et qu'il aura le trait *pluriel*.

(2) Nom[nombre=pluriel genre=\$G]
 " Nom[nombre=singulier genre=\$G]
 Nom_suffixe [nombre=pluriel]

Les règles de production combinent donc les chaînes entre elles. Cette combinaison provoque cependant différents changements orthographiques (suppression de lettres : «nez+s» → *nez*; ajout : «gros+e» → *grosse*; changement : «beau+s» → *beaux*). Ceux-ci sont pris en compte par les règles morphographémiques.

d. L'ensemble de **règles morphographémiques** établit une correspondance entre les chaînes de symboles lexicaux (qui proviennent de la concaténation de mots complets et partiels, par exemple «beau+s» en (3)) et celles de surface (le mot après modification orthographique, comme *beaux*).

forme de surface	b	e	a	u		x
forme lexicale	b	e	a	u	+	s

(3) Tableau

Les règles morphographémiques se présentent comme en (4).

(4) Opérateur Contexte gauche - Focus - Contexte droit Contraintes

Ici, *Opérateur* indique s'il s'agit d'une règle optionnelle ($=>$), de coercion ($<=>$) ou obligatoire ($<=>$); *Focus*, la correspondance entre le niveau lexical et le niveau de surface; *Contexte gauche* et *Contexte droit*, les contextes dans lesquels la correspondance est possible (règle optionnelle), autorisée (règle de coercion) ou obligatoire. Enfin, *Contraintes* est une liste d'ensembles typés de traits qui restreint l'application de la règle à certaines constructions. *Focus*, *Contexte droit* et *Contexte gauche* spécifient des paires de symboles lexicaux et de surface, comme en (5).

(5) symbole de surface / symbole lexical

Prenons comme exemple le cas de la règle qui supprime le morphème «s» du pluriel après un «z» final («nez+s» → *nez*) :

(6): $<=>$ z/z + - $<>/s$ - ~ Plur_suffixe[]

Celle-ci spécifie que le «s» ne correspond à aucun symbole de surface ($<>/s$) si le contexte gauche est un «z» au niveau lexical et de surface (z/z). Cette règle est contrainte par les informations du champs *Contraintes* : elle ne s'appliquera qu'en présence d'affixes de type *Plur_suffixe* (c'est-à-dire les suffixes du pluriel). Le symbole (+) indique le point d'attache des affixes et le tilde (~), la fin du mot.

mmorph facilite la description des règles morphographémiques par l'usage de macros. Par exemple, la règle de (6) s'applique aussi après le «s» et le «x» («prix+s» → *prix*; «repas+s» → *repas*). Plutôt que de définir plusieurs règles, le linguiste peut remplacer le *Contexte gauche* par la macro *s_ou_x_ou_z*, comme en (7a). Celle-ci correspond à la disjonction des trois paires s/s, x/x et z/z, comme l'illustre (7b). Ainsi, la règle s'appliquera après un «s», un «x», ou un «z».

(7) a. $<=>$ s_ou_x_ou_z + $<>/s$ - ~ Plur_suffixe[]
 b. s_ou_x_ou_z : s/s x/x z/z

mmorph est donc un système basé sur règles, qui s'oppose aux listes de mots complets : le fait d'utiliser des règles permet de tirer des généralisations linguistiques. Celles-ci facilitent ainsi considérablement la modification et l'extension de lexiques (Bouillon et Tovenà, 1991).

La suite montre cette flexibilité et le pouvoir de généralisation. Elle décrit d'abord les règles puis les lexiques.

3. LES RÈGLES

3.1 Français

Notre morphologie du français comporte quarante-sept règles de production et sept règles morphographémiques. Le français présente trois particularités, par rapport à une langue comme l'anglais : tout d'abord, elle nécessite souvent le changement de plusieurs lettres à la fois («chameau+e» → *chamelle*); ensuite, l'application des règles ne peut être prédite de la forme lexicale («appel+e» → *appelle*, mais «pel+e» → *pèle*). Enfin, certaines règles sont facultatives. Nous les examinons successivement.

- a. **Mise en correspondance d'une séquence de symboles** : dans beaucoup d'analyseurs morphologiques de ce type (CLE, par exemple, voir Rayner *et al.*, 1996 et Carter, 1995), les règles morphographémiques ne peuvent faire correspondre qu'une paire de symboles à la fois. Pour traiter des cas comme «chameau+e» → *chamelle*, le linguiste doit donc écrire deux règles : l'une qui établit la correspondance entre le «a» et le «l» devant la paire «l/u» et l'autre qui fait correspondre «u» à «l» devant le suffixe du féminin «e» muet. Il s'ensuit une multiplication des règles. Dans *mmorph*, celle-ci peut être évitée (contrairement à ce que soutient Ruessink, 1989), puisqu'une même règle peut porter sur des séquences de symboles. La règle en (8), par exemple, spécifie que les symboles lexicaux «a» et «u» de *chameau* correspondent tous les deux au symbole de surface «l» après un «e» et devant un «e» muet.

(8) <=> e/e - l/a l/u - + e/e_muet

La description des verbes du troisième groupe de conjugaison est ainsi grandement facilitée («peign+ra» → *peindra*; «dev+e» → *doive*, etc.).

- b. **Contrainte sur l'application des règles** : dans *mmorph*, différentes possibilités peuvent être envisagées pour contraindre l'application de règles : tout d'abord, on peut choisir de ne décrire que les règles générales et d'énumérer les exceptions (comme dans Bouillon et Tovenà; 1990, Ostling Anderson, 1987). En français, cependant, une chaîne lexicale peut avoir plusieurs équivalents au niveau de surface, sans que l'on puisse décider quel est le cas général et quelles sont les exceptions. Par exemple, les verbes, les noms et les adjectifs qui se terminent par «el» et «et» peuvent doubler le «l» ou le «t» devant un «e» muet («appel+e» → *appelle*) ou changer le «e» préfinal en «è» («pel+e» → *pèle*). Dans ce cas, quelle est l'exception ? Deux autres solutions peuvent aussi être envisagées en *mmorph* : on peut introduire un

caractère lexical spécial dans les entrées lexicales pour distinguer les différentes chaînes entre elles ou ajouter des contraintes dans les règles morphographémiques. Illustrons ces deux possibilités avec le traitement des mots se terminant en «el». Pour distinguer les mots qui doublent la consonne de ceux qui changent le «e» préfinal en «è», nous remplaçons d'abord le «e» de la racine par le symbole lexical «e_fort» dans le premier cas et nous gardons le «e» dans le second, comme en (9) (dans l'entrée, un symbole lexical composé de plusieurs caractères commence par «&» et se termine par «;» pour éviter la confusion avec les symboles lexicaux qui ont une seule lettre).

- (9) a. "app&e_fort;l" = "appeler"
 b. "pel" = "peler"

Dans la règle morphographémique qui double le «l» (10), nous précisons ensuite que le changement ne se produira qu'en présence de ce symbole dans le contexte gauche. Elle ne pourra donc pas s'appliquer à *peler*.

(10): <=> e/e_fort - <l l>/l - + Fut_e_muet

Le problème est en fait plus complexe pour deux raisons différentes :

- (i) d'une part, le «e» (qui correspond au premier caractère du suffixe du futur et du conditionnel) peut parfois être considéré comme un «e» muet qui affecte les lettres qui le précèdent, sans qu'il s'agisse d'une règle générale. Par exemple, «appel+erai» → *appellerai* (avec doublement du «l» au futur et au conditionnel), mais «céd+erai» → *céderai*, et non pas **cèderai*, comme on s'y attendrait au vu de «céd+e» → *cède*. Pour rendre compte de cette différence, nous utilisons différents symboles lexicaux qui distinguent parmi les affixes de type *Verbe_suffixe* ceux qui ne sont pas muets («ez») (11a), ceux qui le sont («e»). (11b) et ceux qui peuvent être considérés comme tels («erai», «erais») (11c).

- (11) a. "ez" Verbe_suffixe[...]
 b. "&e_muet;" Verbe_suffixe[...]
 c. "&fut_e;rai" Verbe_suffixe[...]

Dans les règles morphographémiques, nous spécifions ensuite si la règle s'applique ou pas devant le "fut_e" du futur et du conditionnel. En (10) par exemple, la règle qui double le «l» s'applique devant le «e» muet, y compris le «e» du futur et du conditionnel, puisque *Fut_e_muet* est une macro qui définit les deux paires de (12).

(12) Fut_e_muet : e/e_muet
 e/fut_e

En revanche, la règle qui change le «é» en «è» ne s'applique que devant le «e» muet, comme exemplifié en (13).

(13) <=> - é/è - Consonne + e/e_muet

- (ii) D'autre part, les noms et les adjectifs peuvent aussi doubler le «l» et le «t» après d'autres voyelles, comme le «i» («gentil+e» → *gentille*) et le «u» («nul+e» → *nulle*). Pour éviter que cette règle ne s'applique aux verbes («pil+e» → **pille*), nous limitons son application au morphème du féminin et nous ajoutons une contrainte dans la règle qui effectue le doublement : (14) ne s'appliquera que devant le morphème «e_muet» qui a le type Fem_suffixe.

(14) <=> u_ou_i - <ll>/l - + e/e_muet **Fem_suffixe** []

- c. **Les règles optionnelles** : certaines règles en français sont optionnelles, comme celle qui change le «y» en «i» si cette lettre suit un «a» et précède un «e» muet («pay+e» → *paie* ou *paye*). Pour ce faire, nous utilisons l'opérateur => qui indique que le changement morphographémique est optionnel (15). Notons que ce changement peut aussi avoir lieu devant le «e» du futur et du conditionnel (Fut_e_muet).

(15) => a/a - i/y - + Fut_e_muet

3.2 Italien

Notre traitement de l'italien se fait par dix règles de production et seize règles morphographémiques. Ici, nous nous intéresserons à la règle d'insertion du «h» et au traitement des clitiques.

- a. **Insertion du «h»** : la plupart des mots dont la racine se termine par «g» et «c» insèrent la semi-consonne «h» devant tout suffixe commençant par «e» ou «i» («elvetic+e» → *elvetiche*, *helvétiques*; «acciu+e» → *acciughe*, *anchois*; «trasloc+i» → «*traslochi*», *tu déménages*). Pour opérer ce changement, nous définissons la règle en (16).

(16) <=> c_ou_g - h/<> ~ + e_ou_i

Les mots qui n'obéissent pas à cette règle reçoivent un caractère lexical spécial à la place du «c» et du «g», comme en (17) pour l'équivalent italien de *faux*.

(17) "fal<no_h;" = "falce"

- b. **Clitiques** : une des particularités de l'italien est la présence de clitiques qui s'attachent à certains modes et temps (impératif, gérondif, participe présent et infinitif) au verbe qui le sous-catégorise, de la manière résumée dans le tableau (18).

Impératif singulier avec les différentes réalisations de clitiques				
	Sans clitique	Clitique accusatif (c1A)	Clitique datif (c1D)	Clitiques accusatif et datif (c1A et c1D)
intransitif	parti ! <i>pars !</i>	-	-	-
transitif direct	vinci ! <i>vaincs !</i>	vincilo ! <i>vaincs-le !</i>	-	-
transitif indirect	obbedisci ! <i>obéis !</i>	-	obbediscigli ! <i>obéis-lui !</i>	-
bi-transitif	dedica ! <i>dédie !</i>	dedicalo ! <i>dédie-le !</i>	dedicami ! <i>dédie-moi !</i>	dedicamelo ! <i>dédie-le-moi !</i>

(18) Tableau

Pour traiter les clitiques et générer toutes ses formes, nous avons besoin des informations suivantes :

- (i) Dans l'entrée lexicale, nous précisons si le verbe prend un clitique accusatif (c1A) ou datif (c1D). «*Dedicare*» (bi-transitif), par exemple, peut avoir un clitique accusatif et datif (la valeur de c1A et c1D est vide), alors que «*vincere*» (transitif) n'a qu'un clitique accusatif (la valeur de c1D est aucun). Enfin, «*partire*» qui est intransitif n'accepte pas de clitique.

- (19) a. Racine_verbe [c1D=vide c1A=vide ...] "dedic" = "dedicare"
 b. Racine_verbe [c1D=aucun c1A=vide ...] "vinc" = "vincere"
 c. Racine_verbe [c1D=aucun c1A=aucun ...] "part" = "partire"

- (ii) Les règles de production combinent le verbe avec les clitiques qu'il admet. Ainsi, deux règles combinent un verbe avec le clitique datif ou accusatif si ceux-ci sont possibles. Une troisième prend pour entrée un verbe (avec le clitique datif) et y ajoute le clitique accusatif, en vérifiant certaines contraintes auxquelles sont soumis les clitiques italiens. Par exemple, la séquence «*lele*» est impossible («**decicalele*», *dédie-le-lui*).
- (iii) Lors de la composition du verbe et du clitique se produisent différents changements orthographiques qui sont pris en compte par les règles orthographiques : celles-ci changent notamment le «*i*» final du clitique datif en «*e*», s'il est suivi d'un des clitiques accusatifs, soit «*dedica+mi+lo*» → *dedicamelo*.

3.3 Allemand

Notre morphologie de l'allemand comporte seize règles de production et cinquante-cinq règles morphographémiques. Nous discutons ici l'«umlautung» (insertion du tréma) et des cas de dérivation et montrons comment ils ont été traités dans *mmorph*.

- a. **Insertion du tréma** : l'insertion du tréma est peut-être le phénomène morphographémique le plus connu de l'allemand. Il concerne le changement systématique de la voyelle de la racine «a», «au», «o», «u» (parfois «e») en «ä», «äu», «ö», «ü» (et «ie»). Celui-ci se produit dans la morphologie flexionnelle («das Haus», la maison → «die Häuser», les maisons) et dérivationnelle («der Tag», le jour → «täglich», quotidien) (Trost, 1990).

Pour le traitement, nous avons recours aux symboles lexicaux «a_tréma», «o_tréma» et «u_tréma» qui marquent les voyelles sujettes à l'insertion du tréma. Comme ce phénomène ne se produit que devant des suffixes spécifiques, nous précisons à l'aide de traits le contexte dans lequel ce changement va avoir lieu. Ainsi pour les noms, la règle (20) indique que ce phénomène n'apparaît que lorsqu'on ajoute à la racine du nom un suffixe du pluriel :

(20) <=> - ä/a_tréma - Nom_suffixe[nombre=pluriel]

Le traitement n'est cependant pas encore complet : comme «a_tréma» ne correspond pas à une lettre de surface, nous devons spécifier comment il se réalise quand il se combine avec les autres suffixes. Au lieu de tous les énumérer dans une règle, nous pouvons exploiter les règles morphographémiques optionnelles, comme en (21). Celles-ci s'appliquent dans tous les cas sauf ceux qui sont traités par une règle obligatoire. Par exemple, lorsque nous ajoutons à un nom un des suffixes du génitif singulier, le symbole lexical «a_tréma» sera changé en «a» («das Haus», la maison → «des Hauses», de la maison).

(21) => - a/a_tréma -

- b. **Morphologie dérivationnelle** : *mmorph* permet également de traiter la morphologie dérivationnelle. Notre morphologie ne traite pour l'instant que les cas suivants : les adjectifs et les infinitifs qui peuvent être nominalisés («rot», rouge → «das Rot», le rouge; «essen», manger → «das Essen», → le manger) et les participes qu'on peut utiliser en tant qu'adjectifs («gelesen», lu → «das gelesene Buch», le livre qui a été lu). L'exemple (22) montre une version simplifiée de la règle qui change un participe (présent, prp ou passé, psp) en un adjectif non fléchi. Dans (22), le signe «|» indique une disjonction : la règle s'applique si la valeur de vform est prp ou psp.

(22) Adjectif[type=non_fléchi degré=pos] Verbe[vform=prp|psp]

4. LEXIQUES

Les descriptions linguistiques existent pour le français, l'italien, l'allemand et l'anglais et sont disponibles sur demande². Ils ont été créés de manière semi-automatique, à partir de dictionnaires existants et ont été validés à l'aide de diverses listes de mots

2 L'outil *mmorph* est disponible à l'URL <http://www.issco.unige.ch/tools/>.

complètes et de correcteurs d'orthographe. Le tableau (23) donne le nombre de mots de base et de formes dérivées.

	Français	Italien	Allemand	Anglais
Nombre de mots de base	22236	29621	45693	25955
Nombre de formes dérivées	235598	365929	3392697	53214

(23) Tableau

Précisons finalement que la description morphosyntaxique se base sur les spécifications développées dans le projet *Multext* (Calzolari et Monachini, 1995).

5. CONCLUSION

Dans cet article, nous avons montré la flexibilité et les possibilités de *mmorph*, à travers les descriptions linguistiques de plusieurs langues. Les lexiques continuent à être développés pour couvrir des domaines plus spécialisés et nous prévoyons plusieurs extensions de l'outil, en particulier pour le traitement des noms composés et les changements morphographémiques externes au mot, très fréquents en français («le++état» → *l'état*; «va++il» → *va-t-il*; etc.).

RÉFÉRENCES

- ARMSTRONG, S. (1996) : «MULTEXT : Multilingual Text Tools and Corpora», H. Feldweg et E. W. Hinrichs, *Lexikon und Text*, Sonderdruck aus *Lexicographica*, Series Maior, Band 73, Max Niemeyer Verlag, pp. 107-119.
- BOUILLON, P. et L. TOVENA (1990) : *L'analyse morphologique du français et de l'italien avec le lexique ALVEY*, ISSCO Working Paper, n° 57.
- BOUILLON, P. et L. TOVENA (1991) : «Word formation and computational dictionaries», *Actes de TKE '90*, pp. 447-454.
- CALZOLARI, N. et M. MONACHINI (1995) : *Common Specifications and Notation for Lexicon Encoding and Preliminary Proposal for the Tagsets*, Deliverable 1.6.1B, MULTEXT.
- CARTER, D. (1995) : «Rapid development of morphological descriptions for full language processing systems», *Actes de ACL*, également sous SRI, Technical Report CRC-047.
- KAPLAN, R. M. et M. KAY (1994) : «Regular Models of Phonological Rule Systems», *Computational Linguistics*, 20, 3, pp. 331-378.
- KARTTUNEN, L. et A. K. WITTENBURG (1983) : «A two-level morphological analysis of English», J. Texas (dir), *Linguistics Forum*, 22, pp. 217-228.
- KOSKENNIEMI, K. (1984) : «A General Computational Model for Word-form Recognition and Production», *Actes de COLING*, pp. 178-181.
- OSTLING ANDERSON, A. (1987) : «L'identification automatique des lexèmes en français contemporain», *Acta Univ. Ups. Studia Romanica Upsulienisa* 39, Uppsala.

- PETITPIERRE, D. et G. RUSSELL (1995) : *MMORPH - The Multext Morphology Program*, Version, 2.3, Deliverable 2.3.1, MULTEXT, ftp://issco-ftp.unige.ch/pub/multext/mmorph.doc.ps.tar.gz.
- PULMAN, S. G. et M. R. HEPPLER (1993) : «A Feature-Based Formalism for Two-Level Phonology: A Description and Implementation», *Computer Speech and Language*, 7, pp. 333-358.
- RAYNER, M., CARTER, D. et P. BOUILLON (1996) : «Adapting the Core Language Engine to French and Spanish», *Actes de la conférence internationale NLP+IA/TAL+AI*, Moncton, NB, Canada, pp. 224-232.
- RITCHIE, G. D., RUSSELL, G. J., BLACK, A. W. and S. G. PULMAN (1992) : *Computational Morphology: Practical Mechanisms for the English Lexicon* Cambridge, MA, The MIT Press.
- RUSSINK, H. (1989) : *Two-Level Formalisms*, Working Papers in Natural Language Processing, n° 5, Rijksuniversiteit Utrecht.
- SHIEBER, S. M. (1986) : *An Introduction to Unification-Based Approaches to Grammar*, Stanford, CSLI.
- THEOFILIDIS, A. et S. RIEDER (1995) : *Final Specifications on German Inflectional Morphology*, Deliverables D2.2 et D3.1, MLAB93-17.
- TROST, H. (1990) : «The Application of Two-level Morphology to Non-Concatenative German Morphology», *Actes de COLING*, pp. 371-376.

COMMENT REPRÉSENTER L'EXPÉRIENCE INDIVIDUELLE QUI DONNE LEUR SENS AUX MOTS, APPROCHE INFORMATIQUE

Françoise FOREST

Groupe «Langage & cognition», LIMSI, Orsay, France

1. LE PROBLÈME DE LA REPRÉSENTATION DU SENS EN LANGUE

Du point de vue de l'informatique, notre travail se situe au carrefour de plusieurs domaines, le traitement automatique des langues, les systèmes de recherche d'information, la reconnaissance des formes, l'apprentissage automatique. Du point de vue des sciences cognitives, il a l'ambition d'étudier les possibilités de modélisation des relations qui se créent entre les mots, les perceptions et les abstractions au cours de l'acquisition et de la représentation des connaissances, de proposer une approche qui associe le perceptif, notamment dans sa dimension topologique et continue, et le conceptuel traditionnellement représenté comme discret. Les choix informatiques étant dictés par l'approche cognitive adoptée, nous commencerons par présenter et justifier cette dernière.

1.1 Le point de vue cognitif

L'hypothèse fondamentale sur laquelle s'appuie ce travail de recherche est l'**existence d'un sujet** dans le processus de compréhension. Il s'agit d'abord pour nous de représenter le sujet pour lequel l'énoncé émis fait sens. L'énoncé étant exprimé au travers d'une langue, nous nous intéressons aux liens que ce sujet a tissés, au cours de son histoire personnelle, entre les mots et les situations perçues, et tout particulièrement aux situations d'apprentissage langagier, celles au cours desquelles le sujet a pour la première fois mis en relation le monde perçu et les mots entendus. Pour développer cet aspect, nous nous inspirons principalement de travaux effectués en psycholinguistique par L.S. Vygotski (1985).

La deuxième hypothèse qui s'est avérée nécessaire est en relation avec la thèse d'un codage multiple des connaissances (Denis, 1989). Toute compréhension se ramenant à une transformation de représentation, il nous est apparu rapidement que la représentation informatique des situations perçues ne pouvait pas se réduire à une description propositionnelle. Nous faisons donc l'hypothèse d'une capacité que nous qualifions de **perception cognitive**. Il s'agit de rendre compte des situations perçues en termes d'entités qui interagissent dans un espace métrique. Cette hypothèse permet de proposer des représentations qui préservent les relations de proximité et de succession présentes dans la

situation. Nous nous inspirons principalement de travaux effectués par R. Thom (1977 et 1991).

La troisième caractéristique de notre travail, d'un point de vue cognitif, est la volonté de s'inspirer largement des particularités de la **langue des signes** dans la représentation de l'expérience acquise par le sujet. Deux aspects de cette langue nous intéressent. Tout d'abord son expression spatiale et temporelle permet de dessiner les situations, et donc d'en fournir simplement une représentation telle que celle souhaitée plus haut. Ensuite la capacité de la langue des signes à «dessiner» aussi simplement des situations abstraites que des situations concrètes suggère que la représentation spatio-temporelle proposée est adaptée à la mise au point de processus de compréhension par analogie qui s'apparentent à ceux qu'on peut imaginer à l'œuvre quand il s'agit d'exprimer des notions abstraites à l'aide de gestes.

1.1.1 Les études sur l'acquisition du langage chez l'enfant, les travaux de Vygotski

Les études sur l'acquisition du langage chez l'enfant mettent en évidence l'importance de la communication langagière avec les adultes au cours de la petite enfance, ainsi que l'importance de l'expérience perceptive et affective vécue par l'enfant.

Pour Vygotski, cette acquisition est la face visible d'une acquisition plus importante, celle des concepts.

Vygotski est un psycholinguiste russe du début du vingtième siècle qui s'est notamment intéressé au problème de l'apprentissage langagier chez l'enfant. Il aborde l'étude de la pensée verbale du point de vue développemental. Dans son ouvrage *Pensée et Langage*, il distingue les concepts «spontanés», qui s'acquièrent au cours de l'expérience quotidienne de l'enfant, expérience soutenue par l'interaction qu'il entretient avec les adultes, et les concepts «scientifiques», qui sont d'abord rencontrés par l'enfant sous forme de définitions verbales. Pour lui, un concept n'existe que dans un réseau de significations. C'est une structure dynamique qui se construit au cours de la résolution d'un problème, et évolue constamment au cours de l'évolution du développement de l'individu.

En ce qui concerne l'acquisition des concepts spontanés, Vygotski met en évidence l'existence intermédiaire d'agglomérats qu'il nomme **complexes**. Ces complexes réunissent un ensemble d'expériences individuelles présentant une certaine similitude aux yeux de l'enfant, et qu'une ou plusieurs caractéristiques perceptives rapprochent. Parmi ces caractéristiques perceptives communes, un rôle très particulier est joué par les mots énoncés par les adultes pour nommer les entités perçues. Vygotski justifie ainsi l'importance primordiale de l'interaction langagière avec les adultes dans l'étape intermédiaire de construction des complexes. Le mot joue dans cette construction un rôle central, puisqu'il est d'abord *objet de perception* avant de devenir également *élément de signification*. C'est sa présence dans le psychisme de l'enfant qui permet à celui-ci d'abstraire une notion de l'ensemble de ses expériences perceptives.

On trouvera dans Forest et Siksou (1994) un exposé de l'intérêt que nous lui trouvons encore aujourd'hui dans une perspective de modélisation du sens.

1.1.2 Un ancrage perceptif de la structuration des connaissances

Les résultats de travaux plus récents sur l'acquisition du langage chez l'enfant montrent qu'un enfant est, même dans son stade de développement préverbal, capable de distinguer les entités en jeu dans une situation qu'il perçoit (Rondal, 1983). Nous ajoutons qu'il est capable d'effectuer un repérage de ces entités dans son champ de vision (plus haut, plus bas, à gauche, à droite, contre, dans...). Cette capacité de reconnaissance perceptive est soutenue par le dialogue que l'enfant poursuit à chaque instant avec son entourage. En effet, on constate que les adultes en interaction avec un très jeune enfant utilisent de nombreux moyens pour l'aider à catégoriser le monde et à agir sur lui (gestes et intonations soutenant le message verbal, actions d'accompagnement, appels à la dimension affective...). Nous parlerons d'une *dimension cognitive de la perception* qui nous paraît ressortir de la même approche que celle utilisée dans des travaux comme ceux de Langacker (1986) sur la grammaire cognitive, de Thom et de Petitot (Petitot, 1991) sur la sémiophysique et la morphodynamique.

Thom appuie son travail sur la notion de discontinuité. Dans le domaine de la perception, il s'agit à la fois de discontinuités spatiales et temporelles. C'est la capacité de reconnaissance des discontinuités qui permet de percevoir les différentes entités en jeu dans une situation perçue. Thom parle à leur propos de formes saillantes dans un espace substrat. Concernant la représentation de cette perception dans le langage, il propose seize morphologies archétypes correspondant à seize verbes d'action ou d'état, et interprétables comme les transformations de relations topologiques entre les entités interagissant dans une situation. On rencontre aussi chez Langacker cette proposition d'une représentation topologique dynamique des objets du discours.

Ces approches ont en commun de proposer une représentation spatiale et dynamique des situations décrites ou observées. Ce type de représentation, associé à des descriptions langagières, nous rapproche de l'hypothèse de double codage (Pavio, Denis). Nous pensons que ce peut être un fondement productif pour une modélisation opérationnelle du phénomène de compréhension. Nous verrons plus loin comment ces approches sont prises en compte dans le modèle que nous proposons.

1.1.3 Les apports de la langue des signes

L'identification des morphologies archétypes proposées par Thom est particulièrement claire dans les langues gestuelles. La langue des signes, tout en disposant, comme les langues orales, d'une panoplie de signes conventionnels (équivalents des mots) pour distinguer les entités du monde perçu, conserve des éléments de nature non discrète servant à construire une représentation spatio-temporelle de la situation décrite dans l'énoncé. Cette représentation rend explicites les relations topologiques entre entités mises en jeu dans cette situation. La capacité de cette langue à exprimer aussi bien des notions abstraites que concrètes suggère une mémorisation spatio-temporelle de tous ces types de situations, les quatre dimensions devenant métaphoriques dans le cas de situations abstraites. L'étude de la forme des énoncés de la langue des signes nous conforte dans notre choix d'une représentation spatiale et temporelle des situations. Nous y voyons la possibilité d'une mise en action simple de processus de compréhension par analogie.

1.2 Le point de vue informatique

1.2.1 Refuser une formalisation a priori des relations de sens

Une des grandes préoccupations du traitement automatique des langues est de mettre en relation des énoncés exprimés en langage naturel et des représentations de ces énoncés traduisant leur signification dans un contexte d'application donné.

Les outils généralement utilisés sont les formalismes logiques manipulant des symboles, ces symboles étant des représentations des mots ou de leur signification (les concepts).

Lorsque les énoncés sont pris dans des textes plus longs que des phrases, ils ont pour rôle de décrire des situations statiques ou dynamiques qu'on se propose de représenter également, par exemple sous forme de schémas, à la Schank. Les structures de données manipulées et les outils logiques pour le faire sont généralement construits de façon ad hoc pour le domaine, généralement restreint, sur lequel ils interviennent.

On a beaucoup reproché à cette approche sa rigidité, son manque de validité sur des domaines plus larges, et les difficultés qu'on rencontre quand on doit faire évoluer les représentations, par exemple les problèmes posés par l'apprentissage de nouvelles connaissances.

1.2.2 Faire l'hypothèse d'une émergence statistique du sens

Les systèmes de recherche documentaire se fondent assez généralement sur cette hypothèse. Ayant pour objectif de retrouver une information stockée sous forme textuelle, dans une base généralement très importante de textes, les outils qu'ils mettent en œuvre sont des outils statistiques. Ils sont robustes et efficaces, en ce sens qu'ils permettent de retrouver en proportion satisfaisante des informations pertinentes malgré le peu de contraintes imposées à la requête de l'utilisateur. Par ailleurs, ils sont souples et permettent de s'adapter assez simplement à un nouveau domaine d'utilisation. En revanche, ne prenant en considération que des fréquences de cooccurrence de mots ou de syntagmes, ils négligent la structure sémantique des connaissances implicitement manipulées. Mais s'ils fonctionnent, c'est sans doute parce que, justement, la cooccurrence des mots dans les textes traduit certaines relations sémantiques. C'est cette idée d'une émergence statistique du sens que nous retenons de cette approche.

1.2.3 Une mise en œuvre de la notion d'expérience individuelle comme fondement de la notion de sens

Les deux approches précédentes ont pour défaut selon nous de se cantonner dans l'étude de données linguistiques. Les travaux effectués en psychologie montrent que la notion de sens, celle de compréhension, ne peuvent se saisir sans faire appel à l'existence du sujet pour lequel cette compréhension a lieu, ce sens se construit. C'est dire que la connaissance ne peut s'appréhender sans une mise en relation de la langue et de l'expérience individuelle, notamment au cours de l'apprentissage du langage. La prise en compte de l'expérience individuelle recouvre, du point de vue de l'informatique, deux types de techniques.

les techniques connexionnistes

L'aspect cumulatif de l'expérience acquise par un individu, les phénomènes de renforcement, d'assimilation, orientent le traitement vers des outils connexionnistes. L'étude des modèles traditionnels (réseaux à couches, cartes de Kohonen et même réseaux récurrents) et des liens qu'ils entretiennent avec les méthodes statiques d'analyse de données nous ont amenée à choisir de développer un réseau à propagation d'activité qui soit exactement adapté à la modélisation des phénomènes décrits dans la littérature psycholinguistique à laquelle nous nous référons (poids de l'affectif dans la mémorisation, importance des liens de cooccurrence entre les mots et leurs situations d'apprentissage, phénomène d'oubli...). Cette structure est détaillée dans le paragraphe de présentation du modèle MoHA. Elle a donné lieu récemment à un travail de thèse effectué par Jean-Pierre Gruselle et à soutenir en octobre prochain.

les techniques de reconnaissance de formes

L'expérience acquise par un individu est d'abord perceptive, généralement centrée sur le visuel. Si l'on dépasse le niveau 2D de perception visuelle, l'acquisition de cette expérience met en jeu des notions de voisinage, d'orientation, d'intensité, qui sont absentes du traitement automatique symbolique classique et qui pourtant entrent en compte dans son ancrage dans le réel perçu de l'expérience individuelle. Plus que la reconnaissance de formes proprement dite, c'est l'aspect topologique des représentations manipulées dans ce domaine de recherche qui me paraît intéressant. Il s'agit de transposer dans le domaine de la représentation de connaissances abstraites les notions de voisinage et de connexité qui permettent de dire que deux formes sont plus ou moins semblables, et de situer les formes les unes par rapport aux autres dans un espace.

Nous pensons qu'une représentation de ce type, utilisable pour les connaissances concrètes visuelles, doit l'être aussi pour les connaissances abstraites. Les travaux menés par les linguistes sur la langue des signes attestent de la possibilité d'une dimension spatiale dans l'expression de toutes les notions accessibles à l'esprit humain (Cuxac, 1993).

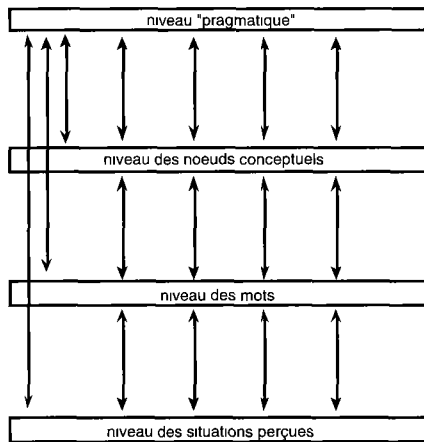
une modélisation opérationnelle sur machine SIMD

Si nous posons qu'une situation perçue est délimitable dans le temps et dans l'espace (Bachelard), que le lien entre le mot perçu et l'objet désigné est clair en situation d'apprentissage (Rondal), que le nombre des échanges langagiers entre enfants et parents est de l'ordre de quelques millions pendant la période d'apprentissage du langage chez le jeune enfant (Lambert et Rondal), une modélisation de la notion de sens nous paraît accessible à l'informatique. Il devient en effet justifié de représenter explicitement les mots, les situations, les occurrences d'entités observées, la taille des corpus autorisant une exploitation réaliste sur les machines massivement parallèles actuelles.

2. UNE PROPOSITION DE MODÉLISATION INFORMATIQUE

2.1 MoHA, Modèle Hybride d'Apprentissage

Le groupe de travail MoHA rassemble au sein du groupe «Langage et cognition» du LIMSI plusieurs chercheurs et doctorants, autour de la construction d'un modèle hybride d'apprentissage. Nous nous situons dans le domaine de l'acquisition de connaissances sémantiques et pragmatiques à partir de données linguistiques et perceptives. Nous faisons l'hypothèse que l'apprentissage ne se réduit pas à des processus ascendants, allant des perceptions vers des connaissances abstraites de différents types, mais qu'il nécessite de mettre en relation l'information nouvellement acquise avec les connaissances déjà structurées, celles-ci guidant le processus d'apprentissage au moyen d'interactions descendantes entre les différents niveaux. Nous faisons également l'hypothèse que l'apprentissage n'est pas un processus séparé des autres processus impliqués dans un système de compréhension, mais qu'il intervient de façon continue, tandis que l'expérience de l'individu s'enrichit. C'est dans l'objectif de modéliser ces phénomènes que nous proposons MoHA, un modèle hybride qui combine une approche numérique et une approche symbolique (Forest et Grau, 1992).



Les différents niveaux de représentation des connaissances et leurs interactions dans le système MoHA

La *première approche* consiste en un traitement massivement parallèle d'événements du monde réel tels que les perçoit un individu. Jusqu'à présent, seule la perception verbale était considérée, ces événements et leur perception verbale étant représentés dans un réseau de type connexionniste, appelé graphe de premier niveau. Il s'agit d'un graphe bipartite qui associe les situations mémorisées aux mots qui en ont accompagné la perception, les liens étant pondérés par des valeurs numériques représentant le poids affectif accordé à cet instant de la mémorisation. Notre objectif est d'en faire émerger des attracteurs qui constituent les concepts «stables». Le rôle du mot dans la formation de ces concepts stables a été étudié, avec l'objectif d'une implantation informatique, dans la thèse de J.P. Gruselle. Les différentes étapes de la construction y

sont spécifiés informatiquement : perception des mots décrivant des situations, mémorisation des situations en mémoire à long terme, émergence de catégories.

La *deuxième approche* est symbolique. Les concepts interviennent dans des informations de plus haut niveau que nous représentons sous forme de schémas. L'apprentissage incrémental de ces schémas s'effectue à l'aide de traitements symboliques tels que l'analogie, la généralisation et la spécification.

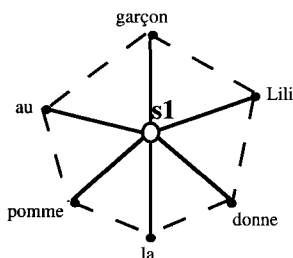
Le modèle MoHA est présenté plus en détail dans un article rédigé en collaboration avec François Bordeaux et Brigitte Grau (Bordeaux, Forest et Grau, 1992).

Nous détaillons ci-dessous les points du modèle MoHA que notre travail actuel aborde plus spécifiquement.

2.2 Le graphe de premier niveau : enrichir la représentation des situations mémorisées

Le graphe de premier niveau rassemble les informations concernant les mots et les situations perçus au cours de l'apprentissage de ces mots.

Chaque situation perçue contribue à ce graphe bipartite qui est constitué de nœuds-mots et de nœuds-situations, chaque nœud-mot étant relié à tous les nœuds-situations représentant des situations à la description desquelles le mot contribue.

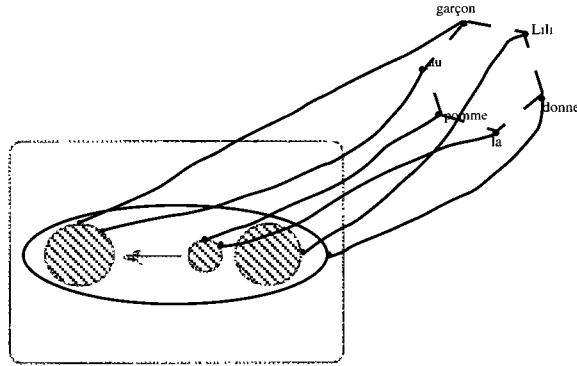


Contribution de la situation s1 décrite par l'énoncé «Lili donne la pomme au garçon» au graphe de premier niveau (1)

Le choix du nombre de mots et de l'importance relative de chaque mot intervenant dans la description ont été particulièrement étudiés par Jean-Pierre Gruselle qui propose plusieurs critères de sélection des mots perçus et différentes lois de diffusion d'activité dans le graphe, simulant le phénomène d'association et permettant notamment de proposer une explication au fait que, dans un contexte donné, un mot n'est généralement pas ambigu.

Pour enrichir la représentation de la situation perçue s1 dans le graphe de premier niveau, il faut substituer au nœud s1 une structure qui fasse apparaître des éléments constitutifs de cette situation, notamment des informations visuelles.

Des travaux ont déjà été réalisés dans notre groupe pour ce qui concerne les situations perçues visuellement. Mais en voulant partir de représentations visuelles en 2D, ces travaux se sont heurtés à des difficultés qui nous éloignaient de notre propos. Un moyen de contourner cet obstacle est de considérer cette modélisation sous l'angle de la morphodynamique, en acceptant d'emblée l'existence d'un niveau cognitif de la perception.



Contribution de la situation s1 au graphe de premier niveau (2) La situation n'est plus représentée uniquement par un nœud dans le graphe de premier niveau, mais par l'ensemble des actants de l'action de «donner» schématisée ici par la flèche

Nous proposons donc de représenter une situation par l'ensemble des actants de l'action décrite, leurs positions relatives dans la scène, et l'évolution de ces positions au cours du temps, notamment l'apparition ou la disparition éventuelle d'une des entités impliquées dans l'action. Nous détaillerons plus loin la forme proposée pour représenter ces informations.

2.3 Le passage au symbolique

Ce passage se manifeste par l'émergence des concepts et l'existence de liens rattachant les situations perçues aux informations conceptuelles et aux schémas du niveau pragmatique.

Si l'on suit l'approche de Vygotski sur la formation des concepts spontanés, le passage le plus difficile à modéliser est certainement le phénomène de «saut qualitatif», celui qui permet à l'enfant de passer de la pensée par complexes à la pensée par concepts.

On a vu comment nous proposons de modéliser les complexes et leur relation aux mots appris. Le passage à une véritable conceptualisation suppose la détection de sous-graphes dont la connectivité interne soit suffisamment forte et les liens aux autres nœuds suffisamment faible pour pouvoir parler d'une certaine «stabilité» du sous-graphe. J.P. Gruselle propose un algorithme de construction des nœuds-complexes associés à ces sous-graphes stables. On peut, dans un premier temps, les assimiler à des nœuds conceptuels, au sens de prototypes d'une catégorie.

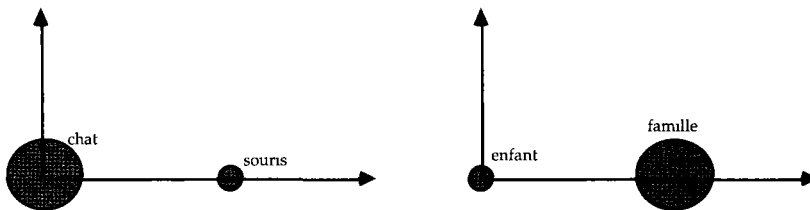
La représentation des entités mémorisées ne se limite pas à des boules dans un espace spatio-temporel. Par l'intermédiaire de l'apprentissage langagier, un certain nombre de caractéristiques verbales sont associées à chaque entité, à commencer par son nom. Si la catégorie abstraite correspondant à la pomme est déjà formée, on associe à la boule un pointeur vers une instance nouvellement créée du prototype de cette catégorie. Sinon, par l'intermédiaire du mot, cette nouvelle entité est mise en relation avec toutes les pommes déjà rencontrées, c'est-à-dire tous les constituants du complexe de pomme.

3. LA REPRÉSENTATION SPATIALE ET DYNAMIQUE DES SITUATIONS

Nous distinguons les *situations statiques*, considérées comme des images fixes, dont l'exploration mentale se fait entité par entité, comme on parcourrait des yeux un paysage, et les *situations dynamiques*, considérées comme une succession de situations statiques à l'intérieur desquelles on aurait la capacité de parcourir transversalement l'évolution d'une entité donnée suivant l'axe temporel. Si on suppose qu'une situation statique est représentée dans un espace à N dimensions, une situation dynamique le sera dans un espace à N+1 dimensions, la dernière représentant l'axe temporel.

3.1 Les situations statiques

Dans une situation statique, les entités présentes sont placées dans un espace métrique à une, deux ou trois dimensions. Ce sont des boules déterminées par leur centre qui indique leur placement dans l'espace, et leur rayon qui peut indiquer leur taille, mais aussi leur saillance dans la situation en fonction d'un certain critère.



Représentation du degré relatif de saillance de deux entités intervenant dans une situation

Un gros *chat* et une petite *souris* n'auront pas le même rayon. Parce qu'ils ne sont pas visuellement de la même taille. Mais aussi parce qu'il est possible de mettre la souris dans le chat, et la possibilité d'établir cette relation topologique suppose que la boule chat puisse inclure la boule souris. Pour des raisons semblables, les entités *enfant* et *famille* seront représentées par des boules de tailles différentes permettant de construire la relation topologique d'inclusion de l'enfant dans la famille.

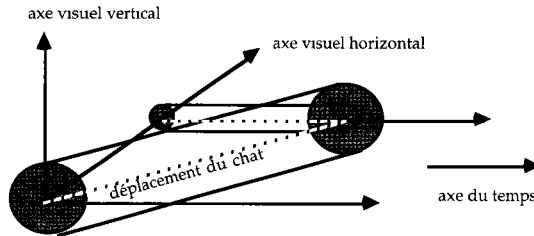
Au démarrage de la construction de MoHA, les premières scènes mémorisées doivent être des scènes concrètes. Mais lorsque le système dispose d'un nombre suffisant de situations concrètes mémorisées, mettant en jeu différentes relations topologiques, on

peut alors, par analogie, construire des représentations de scènes abstraites en utilisant les mêmes techniques. La grande quantité d'expressions et de verbes spatiaux présents dans la langue atteste de la réalité de cette analogie.

3.2 Les situations dynamiques

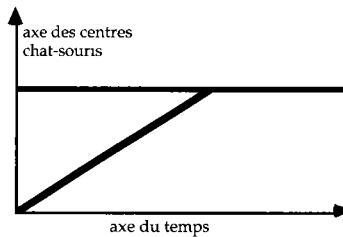
Si les situations statiques peuvent être considérées comme des *instantanés* subjectifs, les situations dynamiques sont des *films* subjectifs. Leur représentation nécessite une dimension supplémentaire qui est celle du temps.

Dans l'espace à N+1 dimensions (N dimensions spatiales et la dimension temporelle), la situation qui évolue est représentée par une forme. Si le chat se précipite sur la souris, la situation dynamique correspondant à la situation statique dessinée plus haut (qui en est le point de départ), aura la représentation donnée ci-après.



Représentation de la situation relative à l'énoncé : «le chat attrape la souris»

En projetant cette forme sur un plan parallèle à l'axe des temps et contenant l'axe des centres chat et souris, on trouve le schéma suivant.

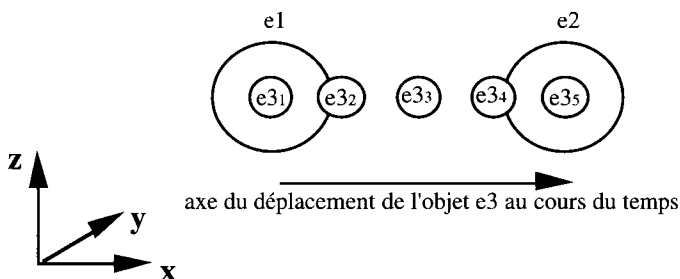


Projection sur le plan (temps, axe des centres chat-souris)

On reconnaît dans ce schéma l'une des morphologies archétypes (Thom, 1977).

3.3 Exemple de représentation dans R5 d'une situation dynamique

On choisit les trois dimensions spatiales, une dimension de prégnance exprimant l'importance relative des différentes entités intervenant dans la situation, et la dimension temporelle.



#donner :

e1 (agent) : $(x=0, y=0, z=0, r=50, t=0)$, $(x=0, y=0, z=0, r=50, t=10)$

e2 (patient) : $(x=200, y=0, z=0, r=50, t=0)$, $(x=200, y=0, z=0, r=50, t=10)$

e3 (objet) : $(x=0, y=0, z=0, r=10, t=0)$, $(x=200, y=0, z=0, r=10, t=10)$

**Représentation de la situation #donner... quelque chose à quelqu'un.
Superposition des plans (x,y) pris à 5 instants successifs de l'évolution
de la situation. e3 se déplace de e1 vers e2**

La dimension de prégnance est subjective. C'est celle qui est caractérisée visuellement par la taille de chaque entité dans l'expression en langue des signes (ici par le rayon des boules). Elle permet d'effectuer des comparaisons analogiques qui se ramènent à des comparaisons de formes, à l'aide de rotations et d'homotéties, dans R5.

3.4 Les situations concrètes reconstruites à partir d'un énoncé

Toute scène intelligible est constituée d'entités qui interagissent. La perception de chaque entité se traduit par la création d'une entité virtuelle dans l'espace mental où se reconstruit cette scène.

Si la scène est visuelle et directement perçue, sa représentation est déduite de la scène réelle. En revanche, si la scène est décrite par un énoncé émis par un locuteur, le travail de l'interlocuteur consiste à reconstituer mentalement la scène à partir des indices linguistiques fournis par l'énoncé et de l'expérience qu'il a préalablement acquise en liaison avec sa propre utilisation de chacun des mots de l'énoncé. Dans ce cas, l'interlocuteur crée une entité virtuelle pour chaque entité dont il est capable de repérer l'existence dans le discours du locuteur, et place cette entité virtuelle dans son espace mental. La place de l'entité virtuelle dans son «champ de vision» mental est soit déterminée par le discours, soit inférée des connaissances préalables. L'existence d'un «champ de vision» suppose que cet espace est construit à partir des perceptions visuelles habituelles. Comme nous l'avons dit plus haut (§1.1.2), nous faisons l'hypothèse que les notions de haut et de bas, de droite et de gauche... y sont déjà présentes au moment de la perception.

Cas où toutes les entités en jeu sont nommées : «La souris est sur l'armoire» fait linguistiquement référence à une scène dans laquelle deux entités entrent en jeu (la souris et l'armoire) dont la relation topologique est déterminée par l'existence de la préposition «sur».

Cas où certaines entités en jeu ne sont pas nommées : «La souris est là-haut» fait linguistiquement référence à la seule souris. «là-haut» détermine sa place en hauteur dans le champ de vision. Cette connaissance peut s'exprimer par le prédicat Haut(souris). Comment en inférer l'existence d'un objet sur lequel la souris devrait se tenir pour que la scène soit acceptable ?

Si, à chaque fois qu'on a eu l'occasion d'observer une souris en hauteur, elle était posée sur un objet, l'examen des situations mémorisées contenant une souris devrait permettre de mettre en évidence la présence obligée de cet objet, même s'il n'en est pas fait mention dans l'énoncé. Le repérage de cette entité non nommée se fait par superposition de toutes les situations concernées, la superposition étant centrée sur l'entité virtuelle associée au mot «souris».

L'observation des relations topologiques qui la concernent dans chaque situation mémorisée permet de mettre en évidence l'existence d'un objet situé dessous, et en contact avec elle. Cela signifie que, dans chacune de ces situations, on trouvera, à la fois sous forme symbolique et sous forme géométrique les informations

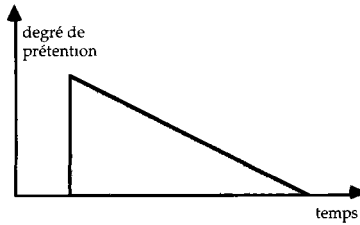
Haut(souris), Sur(souris,x), Contact (souris,x).

La précision de la description de l'entité devant s'identifier à x dans la situation à reconstruire dépend de l'approche qu'on adopte vis-à-vis de la catégorisation. Soit l'entité reste «floue», on ne lui associe que les caractéristiques communes à toutes les observations précédentes, et leur domaine de valeurs possibles. Soit on crée une instance du prototype de la catégorie. Soit on copie à l'identique l'une des instances rencontrées dans une des situations précédemment observées, par exemple celle de la situation la plus prégnante dans le contexte de l'énoncé, c'est-à-dire celle qui est le plus fortement remémorée à la suite de la perception de l'énoncé.

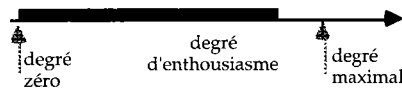
3.5 Les situations abstraites

Même abstraite, une situation concerne toujours des entités qui interagissent. Mais l'interprétation des dimensions spatiales dans lesquelles cette situation est représentée est métaphorique. Chacune des dimensions vaut pour un trait, commun aux entités présentes, et qui ordonne ces entités suivant un ordre total. L'axe des temps demeure en général, même si les valeurs prises par les différentes entités y sont subjectives. Tous les autres axes peuvent être le lieu de relations topologiques entre les entités.

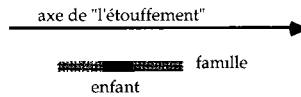
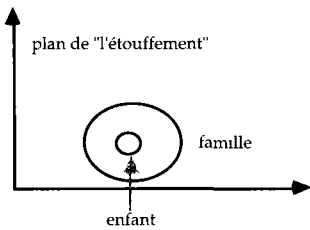
Comment représenter l'expérience individuelle qui donne leur sens aux mots, approche informatique



«une prétention qui diminue»

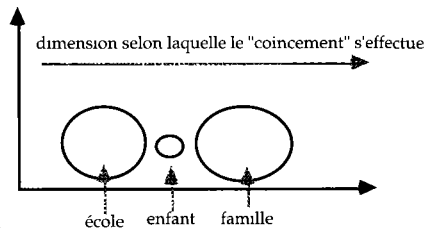


«un enthousiasme fantastique»



«l'enfant étouffé par sa famille»

La relation topologique perçue est la relation d'inclusion.



«l'enfant coincé entre l'école et sa famille»

Les volumes représentent la saillance de chaque entité dans la représentation que construit l'interlocuteur.

4. CONCLUSION

Cette représentation spatiale et temporelle des situations enrichit le graphe de premier niveau. Elle permet d'accéder aux éléments constitutifs de chaque situation dans un format non propositionnel qu'on peut assimiler à un codage de nature différente du codage linguistique utilisé dans le niveau symbolique.

La dimension continue de cette représentation et son caractère topologique permettent d'effectuer des opérations de comparaison, des mesures de proximité interdites dans l'autre type de codage. Elles permettent aussi de mettre en évidence des similarités entre situations qui sont de l'ordre de l'«analogie d'évolution». En assimilant les situations abstraites à des situations concrètes à dimensions effectivement spatiales, on se donne la possibilité d'anticiper l'évolution d'une situation abstraite par comparaison avec des situations concrètes. Dans le cadre d'une approche développementale des connaissances acquises, on voit tout l'intérêt de cette assimilation de l'abstrait au concret, les psycholinguistes ayant généralement observé que l'acquisition des mots désignant des entités concrètes, puis les relations qui les lient, précèdent chez l'enfant l'acquisition des connaissances abstraites.

RÉFÉRENCES

- BORDEAUX, F., FOREST, F. et B. GRAU (1992) : «MoHA, an hybrid learning model: a model based on the perception of the environment by an individual», *IPMU'92-Advanced Methods in Artificial Intelligence*, B. Bouchon-Meunier, L. Valverde, R. R. Yager (Eds), *Lecture Notes in Computer Science*, n° 682, Springer-Verlag.
- CUXAC, C. (1993) : *L'icônicité dans la langue des signes*, VENACO.
- DENIS, M. (1989) : *Images et cognition*, Paris, PUF.
- FERRET, O. et B. GRAU (1996) : «Construire une mémoire épisodique à partir de textes : pourquoi et comment ?», *RFIA'96-Rennes*, 16-18 janvier 1996.
- FOREST, F. (1990) : «Le sens d'un énoncé est fondamentalement lié à l'expérience de l'individu qui le perçoit», *4ème colloque de l'ARC - Paris*, 28-30 mars 1990.
- FOREST, F. (1991) : «Se donner les moyens d'une approche constructiviste de la représentation du sens, le traitement massivement parallèle des données», *Notes et documents du LIMSI*, n° 91-21, 12-1991.
- FOREST, F. et M. SIKSOU (1994) : «Développement de concepts et programmation du sens, Pensée et Langage chez Vygotski», *Intellectica*, vol 1, n° 18, pp. 213-236.
- LAMBERT et RONDAL (1979) : *Le mongolisme*, Bruxelles, Pierre Mardaga éditeur.
- LANGACKER, R. (1986) : «An introduction to cognitive grammar», *Cognitive Science*, n° 10, pp. 1-40.
- PETITOT, Jean (1991) : «Syntaxe topologique et grammaire cognitive», *L'objet, sens et réalité, Langages*, n° 103, septembre.

*Comment représenter l'expérience individuelle qui donne leur sens aux mots,
approche informatique*

- RIVIERE, A. (1990) : *La psychologie de Vygotsky*, Bruxelles, Mardaga.
- RONDAL, Jean-A. (1983) : *L'interaction adulte-enfant et la construction du langage*, Bruxelles, Mardaga.
- THOM, René (1977) : *Stabilité structurelle et morphogénèse*, 2^e édition, Paris, Interéditions.
- THOM, René (1991) : *Esquisse d'une sémiophysique*, Paris, Interéditions. 2^e tirage corrigé.
- VYGOTSKI, L.S. (1985) : *Pensée et langage*, Paris, Éditions sociales (traduction de Myschlenie y rech' écrit en 1933 et publié dans *Izbrannye psikhologicheshie issledovanya* en 1956 à Moscou).

VERS L'UTILISATION DES MÉMOIRES DE TRADUCTION POUR LA LOCALISATION DES LOGICIELS : LE PROJET ALAMET

Boubaker MEDDEB HAMROUNI^(1et2), Laurent FISCHER⁽¹⁾ et Mathieu
LAFOURCADE⁽¹⁾

(1) *GETA - CLIPS - IMAG Campus*, (2) *WinSoft SA., Grenoble, France*

INTRODUCTION

Tous les indicateurs économiques montrent la naissance d'une nouvelle phase de traitement de l'information et de la communication : le multilinguisme. Les échanges commerciaux entre les pays évoluent et les moyens informatiques doivent suivre cette évolution. Les constructeurs et éditeurs des outils informatiques avaient jusqu'à présent vécu, plus ou moins tranquillement, des marchés des pays développés. Aujourd'hui, ce n'est plus vrai, et ces constructeurs doivent viser d'autres marchés potentiels.

Cet article analyse les problèmes théoriques et pratiques — aussi bien terminologiques qu'informatiques — liées à la localisation de logiciels et décrit l'architecture d'un système d'Aide à la Localisation de logiciels pAr MEmoire de Traduction (ALAMET). Cet environnement permet d'automatiser la chaîne de traduction d'un produit d'une langue source vers une nouvelle langue cible. ALAMET constitue aussi une solution à la fois technique et théorique à la normalisation de la terminologie utilisée dans les applications informatiques.

Bien que notre approche n'ait pas été complètement implémentée, nous avons pu l'expérimenter partiellement sur un produit anglais et nous avons réussi à obtenir rapidement un produit localisé en deux langues, sans avoir des connaissances dans la langue cible et surtout en un temps beaucoup plus court qu'avec une approche classique manuelle.

1. LOCALISATION DE LOGICIELS ET MÉMOIRE DE TRADUCTION

La localisation d'un logiciel est l'élaboration d'une version dans une langue spécifique adaptée à un nouveau marché. Cette définition reste très grossière, car la localisation va au-delà de la simple traduction des ressources textuelles contenues dans un

programme. Les aspects culturels peuvent avoir un impact sur certains aspects de l'interface utilisateur. Cependant, nous nous limitons ici aux éléments textuels.

D'une façon générale, la localisation d'un produit nécessite deux étapes distinctes :

- une modification profonde du code de l'application source pour permettre la saisie, l'affichage, l'impression et l'échange des documents créés. Ce qui nécessite principalement des compétences techniques;
- La traduction de l'interface utilisateur (alertes, menus, boîtes...), des manuels, etc. Ce qui nécessite principalement une connaissance de la langue cible et une maîtrise d'un certain nombre d'outils informatiques.

Actuellement, le travail de traduction de l'interface utilisateur se fait 1) d'une manière essentiellement manuelle, 2) sans aucune réutilisation des ressources terminologiques qui pourtant existent dans les applications déjà localisées, et 3) sans aide vis-à-vis des problèmes terminologiques.

La mémoire de traduction est une technique qui vise à accélérer le processus de traduction en faisant intervenir des bases de données contenant les traductions précédemment réalisées. De telles bases s'enrichissent au fur et à mesure des traductions effectuées. Elles sont multilingues et peuvent donc faciliter la traduction d'une langue source vers plusieurs langues cibles simultanément. De plus, elles permettent d'augmenter la cohérence terminologique entre produits et entre traductions.

2. OUTILS EXISTANTS

Dans le cadre de l'internationalisation croissante du marché logiciel, les différentes sociétés éditrices de systèmes d'exploitation (Microsoft, Apple, SUN...) et/ou d'applications bureautiques (Microsoft, Adobe...) sont confrontées à la nécessité d'adapter leurs produits à des marchés extérieurs. C'est à ce prix que de telles sociétés peuvent être présentes sur des marchés en pleine expansion (Asie, Moyen-Orient, Europe de l'Est). Si certaines sociétés ne disposant pas des compétences nécessaires (aussi bien linguistiques que techniques) ont été amenées à sous-traiter ce processus, d'autres ont mis en place des outils leur permettant d'automatiser le processus de localisation de leurs produits.

Plusieurs systèmes d'aide à la localisation ont été développés. En général, ils demeurent spécifiques à leur système d'exploitation d'origine et ne sont pas conçus pour traiter une même application portée sur deux plateformes différentes. On peut citer par exemple :

- Le système I.L.O. (Internationalisation et Localisation de l'Offre) de Bull (1996) qui permet la localisation systématique à partir d'une langue source de tout produit déjà internationalisé dans un certain nombre de langues. Cet outil utilise non pas une simple base terminologique, mais un système de traduction

automatique. Une révision humaine est toujours nécessaire avant la génération du produit localisé.

- Dans le monde Unix, un module (IMT) s'intègre au système d'exploitation et offre des capacités multilingues (saisie, affichage) (Archibald & Darisse 1981). Dans le même domaine, on trouve en outre des utilitaires DVX (Development System Extensions) d'aide à la localisation (extraction des chaînes à partir des applications, outils d'indexation).
- Microsoft propose un ensemble d'outils (RLTools — Resource Localization Tools) qui permettent de localiser des applications Windows (3.11, 95, NT). Ces outils ont été utilisés pour la localisation de toute la gamme des produits Microsoft (Word, Excel, Work...) dans leurs versions multilingues.
- Apple propose une application MacOS (AppleGlott) qui permet de localiser des applications Macintosh d'une manière quasi-automatique et sans l'utilisation d'un éditeur de ressources.

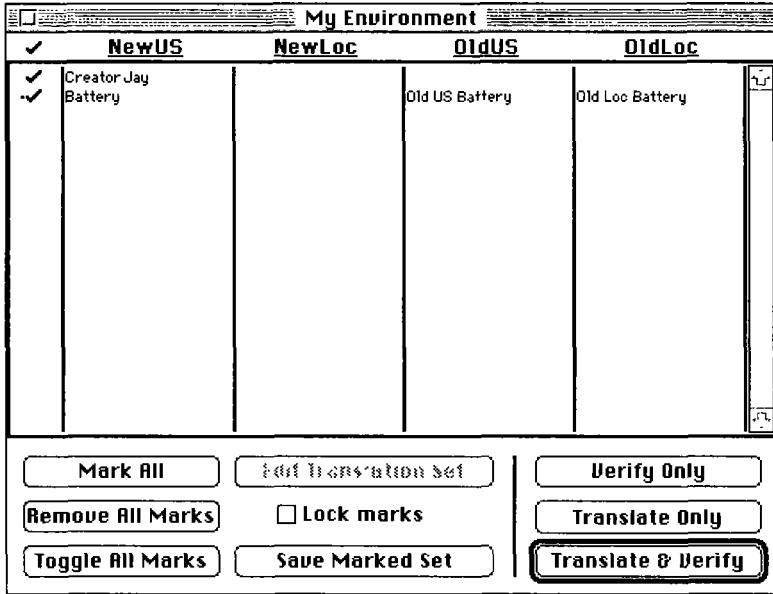
Les outils AppleGlott et RLTools nous ont servi de terrain d'expérimentation.

2.1 AppleGlott (Apple Computer Inc.)

AppleGlott permet de localiser un produit ou un ensemble d'applications Macintosh. Cette application permet non seulement de localiser une application mais donne la possibilité de mettre à jour une application déjà localisée, s'il y a eu de nouvelles versions, en se basant sur un fichier d'historique.

La procédure de localisation commence par la création d'un dossier dans lequel l'application dans la langue source doit être placée. Après le démarrage d'AppleGlott, l'écran suivant apparaît :

- NewUS : contient l'application en langue source (par défaut on suppose qu'elle est en anglais).
- NewLoc : contiendra l'application localisée dans la langue cible.
- OldUS : sert pour une mise à jour et contient l'ancienne version de l'application en langue source.
- OldLoc : contient l'application dans sa précédente version déjà localisée dans la langue cible.



Le bouton «Translate Only», permet d'extraire toutes les chaînes de l'application source. Ces ressources textuelles sont représentées dans le format suivant :

```
{Le texte ici entre accolades est un commentaire}
DITL 128 dialog item text 8 (1) {N'a pas trouvé de traduction}
<File>
<> {entrer la traduction «Fichier» ici }

DITL 128 dialog item text 26 (3) [?? Text Match - Internal Guess]
<Print>
<Imprimer>
{ici il a trouvé en se basant sur une traduction déjà faite dans une version antérieure}
DITL 128 dialog item text 62 (7)
<Settings>
<>

DITL 129 dialog item text 26 (3) [?:.. Wild Guess by position]
<Paper>
< >

STR 128 text 1
<Some Text>
<>
```

Il est nécessaire de saisir manuellement les traductions dans ce fichier texte au niveau de la seconde ligne (entre < et >) et on rappelle le dialogue. «Translate & Verify» permet alors d'injecter les traductions et de générer l'application cible.

L'atout d'AppleGlott est sa simplicité d'utilisation, sa capacité à extraire toutes les chaînes, même si elles ne font pas partie des ressources standards du Macintosh (Str#, MENU...). En plus, la possibilité de mettre à jour une application déjà localisée constitue une fonctionnalité importante. En fait, cet outil gère d'une manière indirecte et implicite une mémoire de traduction.

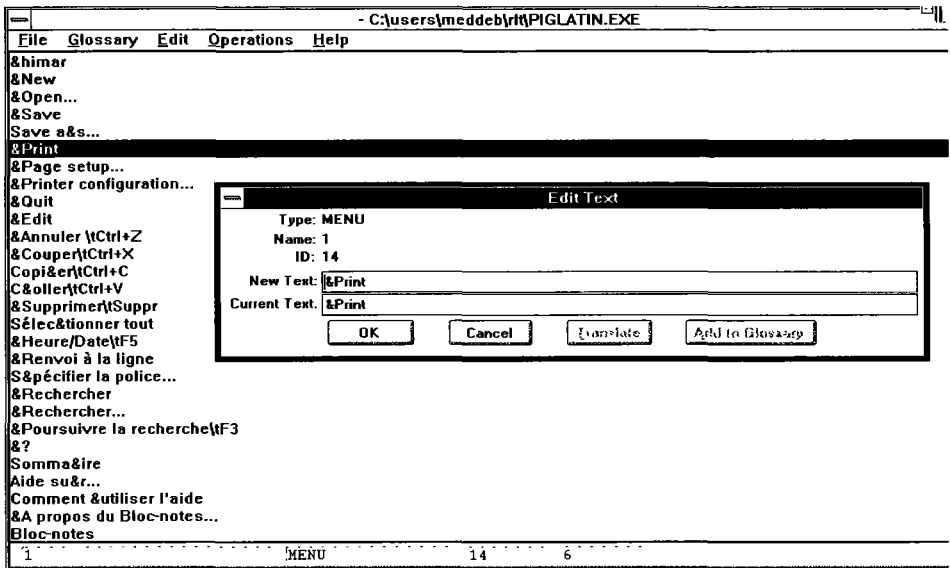
Par contre, tout comme les outils Microsoft, il n'est pas possible de faire les traductions d'une manière automatique, c'est-à-dire qu'on ne peut pas intégrer une mémoire de traduction à l'application pour faire une traduction automatique quand il s'agit de localiser une nouvelle application.

2.2 RL Tools (Microsoft Corp.)

Les outils proposés par Microsoft peuvent localiser un produit, un produit et ses accessoires (filtres, aide...) ou un ensemble de produits. L'outil central (RLQuikEd) existe sous forme d'une application GUI (Graphic User Interface) ou d'un ensemble de commandes DOS. RLQuikEd permet de charger une application et d'en extraire toutes les chaînes. Il propose alors à l'utilisateur une table dans laquelle il doit rentrer la traduction d'une chaîne en regard de la chaîne proposée dans la langue source. Une fois toutes les chaînes traduites, une opération de génération de l'application cible peut être lancée afin d'obtenir le produit localisé.

D'autres outils permettant de réajuster la taille des dialogues, de créer des projets de localisation (nécessaire lors de localisation de plusieurs applications), sont également proposés dans le kit RLTools.

RLTools n'est en fait, qu'un outil d'extraction des chaînes et d'injection du résultat de la traduction. Aucune mémoire de traduction n'est disponible. Il est possible de créer et de réutiliser une mémoire de traduction dans un format texte simple (entrée en langue source, tabulation, entrée en langue cible). Néanmoins, il n'existe pas d'outils de gestion de la mémoire de traduction et aucune possibilité n'est offerte pour suggérer des mots voisins. Par exemple, si on cherche à traduire la chaîne «File:» et si dans la base on a le couple «(File, Fichier)», le système est incapable de fournir la traduction à cause du caractère «:».



2.3 Analyse

Au vu des deux outils présentés ci-dessus, on peut faire un certain nombre de remarques :

- Ces deux outils constituent de bons extracteurs des chaînes à traduire et de bons injecteurs des chaînes déjà traduites.
- Ils n'utilisent pas de mémoires de traduction externes, ou pas efficacement.
- Ils ne permettent pas de faire d'autres actions de localisation (symétrisation des dialogues, réajustement des boutons...).
- Quand ils utilisent des mémoires de traduction, la recherche des chaînes n'est pas très puissante. On aimerait traduire «O&pen» par «O&uvrir» si le dictionnaire contient «Open» et «Ouvrir».
- Ils sont mono-plateforme. Une application Macintosh, bien qu'identique au niveau de son interface sur Windows, se localise avec AppleGlot sur Macintosh et RLTools sur Windows.

Ce dernier point est particulièrement critique. En effet, il est souhaitable de disposer d'un environnement de localisation dont le fonctionnement n'est pas lié à un système d'exploitation ou à un constructeur particulier. D'une façon générale, seule la phase d'extraction et de réinsertion des ressources textuelles reste fortement dépendante de l'origine des produits à localiser.

3. LE PROJET ALAMET

Le projet vise à définir d'une part une architecture générale d'un système de localisation générique, et d'autre part un format de fichier (nommé TRADE) permettant de représenter les unités textuelles en entrée et en sortie de la mémoire de traduction.

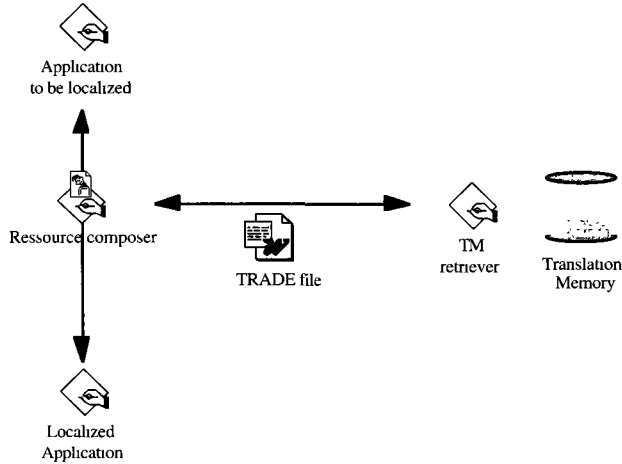
3.1 Architecture

L'architecture du système ALAMET est centrée autour du format de fichier TRADE (dont une description est donnée ci-après). Nous considérons deux principaux composants dans l'architecture ALAMET : 1) le compositeur de ressources et 2) le gestionnaire de mémoire de traduction.

Si on dispose déjà d'une mémoire de traduction, le processus de localisation se décompose de la façon suivante :

- extraire les ressources textuelles de leurs applications sous la forme d'un (ou plusieurs) fichier(s) TRADE (nommé TRADE.in);
- «nourrir» le gestionnaire de mémoire de traduction avec ce fichier;
- obtenir un fichier TRADE (TRADE.out) contenant les ressources lexicales avec leur(s) équivalent(s);
- réviser le fichier TRADE.out. Ceci implique d'une part de vérifier les équivalents proposés, d'en sélectionner un seul pour chaque ressource textuelle, et de compléter ceux qui n'auraient pas été trouvés. Nommons ce fichier TRADE.rev;
- injecter le fichier TRADE.rev dans le compositeur de ressources afin d'obtenir les applications localisées dans les langues cibles;
- injecter le fichier TRADE.rev dans le gestionnaire de mémoire de traduction afin d'augmenter la mémoire de traduction (cette étape est optionnelle, mais néanmoins cruciale pour l'amélioration de l'efficacité du système).

Si le système est vierge de toute mémoire, l'étape consistant à trouver les équivalents doit être faite à la main. Toutefois, en pratique, on cherchera à alimenter dès le départ la mémoire de traduction avec des couples d'applications déjà localisées.



Le compositeur de ressources

Le compositeur de ressources permet l'extraction et la réinsertion des ressources textuelles. Précisément, ce composant doit être capable :

- d'extraire les ressources textuelles d'une application et d'en produire un fichier TRADE;
- de produire une application localisée, à partir d'un fichier TRADE disposant des traductions complétées.

Le processus d'extraction des ressources textuelles dépendant fortement du système d'exploitation et/ou du développeur de l'application, le compositeur prévoit un mécanisme d'extension («plug-ins»). Ces extensions permettent de décrire ce processus d'extraction selon les applications.

Le gestionnaire de mémoire de traduction (GMT)

La tâche majeure du gestionnaire de mémoire de traduction est de compléter un fichier TRADE.in avec les équivalents trouvés dans la mémoire de traduction.

Dans le premier prototype d'ALAMET (en cours de développement), la mémoire de traduction ne propose qu'un mode de correspondance exacte pour la recherche des chaînes. Nous cherchons à intégrer des modes de recherche plus «flous» incluant :

- la gestion de mots-clés (origine de l'application, version, etc.) avec des niveaux de priorité;
- la recherche de chaînes en correspondances partielles.

Le GMT accepte de recevoir en entrée un fichier TRADE en vue de l'augmentation de la mémoire de traduction. De plus, nous envisageons d'accepter des fichiers de dictionnaires terminologiques comme ceux développés dans le cadre des projets FE*.

3.2 Le format de fichier TRADE

Le format TRADE permet de décrire un ensemble de ressources textuelles (en langue source) associées à leur(s) équivalent(s) dans plusieurs langues cibles. Ce format inclut les informations suivantes :

- description globale de l'origine des ressources textuelles (système d'exploitation, applications, version, etc.);
- informations permettant la réinsertion des ressources textuelles en langage cible dans les applications en versions localisées;
- spécification des codages ou des transcriptions utilisés pour les contenus textuels.

De plus, ce format est conçu de façon à être suffisamment lisible par les personnes en charge de la localisation.

Informations d'extractions

Tous les outils existant actuellement dépendent directement des systèmes d'exploitation pour lesquels ils ont été conçus. Nous souhaitons adopter une démarche générique qui rende le composeur de ressources complètement indépendant des systèmes d'exploitation et des éditeurs.

Les informations d'extraction permettent au composeur de ressources de réinsérer les ressources textuelles dans les applications (afin de produire les versions localisées). Il s'agit d'une simple chaîne de caractères. Les informations contenues dépendent directement de l'application à localiser.

Par exemple, sous MacOS on pourra avoir :

```
<Rsc-descriptor>  
DITL 128 dialog item text 62  
</Rsc-descriptor>
```

Les extensions du composeur permettent de générer ces informations en fonction du type d'application concernée. Elles gèrent également le processus inverse de réinsertion de ressources en fonction de ces informations.

Informations textuelles

Les informations textuelles concernent plus directement la chaîne de caractères à localiser. En plus de la chaîne elle-même, nous considérons les informations de formatage et de codage/transcription.

Le formatage gère l'apparence du texte (gras, italique, police, etc.). Sous Windows, le raccourci clavier est indiqué à l'aide du soulignement d'une lettre particulière. Par exemple dans le menu d'une application Windows, on trouve l'article «File» qui permet d'activer la commande d'ouverture d'un fichier en appuyant sur la touche Alt+F. Il est nécessaire de garder ce type d'information lors de la localisation de l'application.

Le code langue indique la langue ou la famille des langues utilisées. En pratique, cette information est rarement explicitée dans les ressources textuelles des applications. Cependant, il est possible d'inclure cette information comme un des paramètres d'extraction.

Le codage indique le type de codage de la chaîne de caractères (par exemple MacRoman, Unicode, ASCII, ASMO449, etc.).

La transcription indique le type de transcription utilisée pour la représentation des chaînes. Cette information est optionnelle.

Informations générales

Les informations générales sont représentées sous forme de variables à valeur ensembliste (énumérées) indiquant des informations comme le système d'exploitation, l'application, etc. Certaines de ces informations peuvent être indiquées globalement. Toutes peuvent être locales à chaque élément textuel.

Nous avons par exemple :

- OS : MacOS, Win95, Win311, ...
- Application : Pagemaker, MSWord, ...
- Version : 1.0, 2.1, ...
- Type : alerte, dialogue, menu

D'autres types de variables peuvent être envisagés.

Le gestionnaire de mémoire de traduction essaye de trouver l'équivalent qui satisfait au mieux les contraintes qu'imposent ces mots clés.

Exemple

```
<TRADE>
<HEAD>
<Source-language>English<\Source-language>
<Coding>MacRoman<\Coding>
<Application> Eudora <\Application>
<\HEAD>
<TRADE-item>
    <Rsc-descriptor>DITL 128 dialog item text 62<\Rsc-descriptor>
    <Source>
    <String>Don't trash<\String>
    <Application>Eudora<\Application>
    <OS>MacOS<\OS>
    <Version>2<\Version>
    </Source>
    <Targets>
        <Target-language>French<\Target-language>
        <String>Ne pas jeter<\String>
        <String>Ne pas détruire<\String>
        <Target-language>Arabic<\Target-language>
        ...
    <\Targets>
<\TRADE-item>
<TRADE-item>
    <Rsc-descriptor>DITL 128 dialog item text 63<\Rsc-descriptor>
    <Source>
    <string>Cancel<\string>
    <\Source>
    <Targets>
        <Application>Eudora<\Application>
        ...
    <\Targets>
<\TRADE-item>
<\TRADE>
```

Comme dans le format HTML, les retours chariot ne sont pas significatifs. Les blancs, tabulations, etc. entre les étiquettes ne sont pas significatifs.

4. ALAMET — UNE PREMIÈRE EXPÉRIENCE

Dans le cadre des projets de dictionnaires Français-Anglais-Malais et Français-Anglais-Thaï soutenus par le MAE, le GETA a acquis une expérience dans la réalisation et la mise à disposition de dictionnaires multilingues (une langue cible vers plusieurs langues sources). Ces projets ont permis d'arriver à une méthodologie efficace de constitution de tels dictionnaires en utilisant des logiciels du commerce.

La société WinSoft a depuis longtemps une expérience dans les logiciels multilingues et dans le service de localisation. Ses clients sont, entre autres, de grandes sociétés de logiciels comme Adobe, Microsoft ou Apple. L'apport de WinSoft consiste essentiellement en deux points :

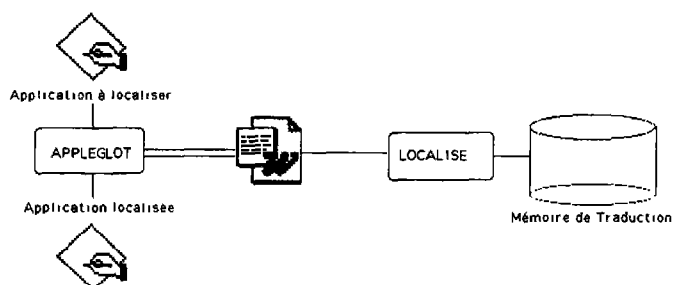
- modifier les logiciels d'origine, pour leur permettre de fonctionner dans des langues utilisant des systèmes d'écriture non romains;
- traduire les ressources de ces logiciels afin de produire des applications localisées pour des marchés comme ceux du Moyen-Orient ou de l'Europe de l'Est.

WinSoft reçoit une demande croissante pour la localisation de logiciels dans des délais courts. Pour ce faire, nous avons conçu une première chaîne de localisation, inspirée de l'architecture d'ALAMET. Nous avons implémenté sous MacOS une première maquette et l'avons expérimenté sur un nouveau produit (PhotoShop 4.0).

4.1 Description

Nous avons utilisé comme outil d'extraction et d'injection le logiciel AppleGlot 2.4 (Apple Computer) et comme outil de gestion de la mémoire de traduction l'application 4D 5.0.2 (ACI).

Notre chaîne de localisation se présente comme suit :



Pour localiser une application, on procède en trois étapes :

1) Utiliser AppleGlot sur l'application source. C'est l'étape d'extraction des chaînes. Cette étape produit un fichier «TEXT» contenant toutes les chaînes à traduire. Ce fichier est dans le format suivant :

```
DITL 128 dialog item text 62 (7)
<Paramètres>
<>

DITL 129 dialog item text 26 (3) [?:•• Wild Guess by position]
<Papier>
< >
```

Les chaînes à traduire sont placées entre les signes «<» et «>» et les chaînes traduites devront être placées dans le «< >» qui suit.

2) Utiliser l'application LOCALISE que nous avons développée. Elle prend en entrée un fichier produit par AppleGlot et cherche les traductions dans la base 4D. Le résultat est un fichier «APPLEGLOT» dans lequel ont été placées les traductions (si elles existent dans la base).

```
DITL 128 dialog item text 62 (7)
<Setting>
<Paramètres >

DITL 129 dialog item text 26 (3) [?:•• Wild Guess by position]
<Paper>
<Papier>
```

L'application LOCALISE est présentée ci-après.

3) Ré-exécuter «APPLEGLOT» sur le fichier produit à l'étape 2 et sur l'application source. C'est l'étape d'injection des ressources. Cela a pour effet de remplacer les chaînes de l'application en langage source par leurs équivalents dans la base, au regard du fichier des traductions et de générer l'application cible dans la nouvelle langue.

4.2 L'outil LOCALISE

LOCALISE a été développé en vue de deux objectifs :

- gérer une base de données terminologiques contenant les termes informatiques les plus utilisés dans les applications Macintosh.

- utiliser la mémoire de traduction pour localiser de nouvelles applications ou mettre à jour des applications déjà localisées.

La base que nous avons construite contient les entrées en 6 langues (français, anglais, arabe, hébreu, japonais, russe) et contient les termes informatiques les plus utilisés dans les interfaces des applications Macintosh.

Cet outil a été écrit avec le langage de programmation de 4D (v5.0.2). Les différentes fonctionnalités relèvent de l'utilisation ou de l'administration :

Utilisation :

Localise permet de lancer la localisation d'une application à travers son fichier de chaînes (glossaire) préalablement produit par AppleGlot. Au lancement de l'action «Localise», l'écran suivant apparaît. L'utilisateur y spécifie :

- 1) le fichier glossaire qui contient les termes dans la langue source,
- 2) le fichier dans lequel se trouveront les chaînes résiduelles qui ne seront pas traduites (car elles sont inexistantes dans la base) et qui seront traduites par un traducteur humain,
- 3) la langue source et la langue cible, puisque la base contient des termes multilingues.

Deux options sont enfin proposées :

- a) chercher des chaînes avec ou sans le caractère «:»;
- b) chercher des chaînes avec ou sans le caractère «...»¹.

Ces deux dernières opérations permettent de trouver le maximum de chaînes.

¹ Le «...» est un caractère ayant son propre codage sur Macintosh.

Construction du glossaire bilingue

Choisir le fichier glossaire (AppleGlot) : **Sélectionner**
Fichier : Dunia:Exemple.txt

Choisir le fichier des chaînes (à traduire) : **Sélectionner**
Fichier : Dunia:Exemple.resultat

Langues :
Source : **Frçais** Cible : **Arabe**

Options :

Chercher les chaînes avec et sans le ":"
 Chercher les chaînes avec et sans le ":"

Annuler **Lancer**

À la fin de l'opération, un nouveau fichier (ayant le même nom mais avec une extension) est produit et contient les termes traduits.

Consulter la base permet d'afficher le contenu de la base selon un ordre alphabétique spécifié via un pop-up menu.

Rechercher permet de rechercher des termes dans la base.

Administration :

Les fonctionnalités proposées permettent de supprimer, ajouter, modifier, importer, exporter des termes. Ce sont les opérations classiques qu'on trouve dans la plupart des gestionnaires de bases de données.

4.3 Résultats

Cette chaîne de traduction a été utilisée pour localiser vers l'arabe la nouvelle version de Photoshop à partir d'une version préalablement localisée, Photoshop 3.0.5. Grâce à notre méthodologie, nous avons réussi d'une part à récupérer toute la terminologie de la précédente version d'une manière automatique, et d'autre part à trouver environ 30 % des nouvelles chaînes dans notre mémoire de traduction. C'est un résultat intéressant, puisque les nouveaux termes sont ceux inhérents au domaine du traitement d'images, ce qui donne a priori peu de chances de les trouver dans notre base.

Il est important aussi de noter que l'opération de localisation a pris beaucoup moins de temps qu'une approche manuelle (déjà réalisée par la société qui édite le logiciel, Adobe) puisque nous avons réussi à récupérer automatiquement les chaînes existantes dans la précédente version et 30 % des nouvelles chaînes lors du temps d'exécution du programme d'injection, soit environ une heure.

Enfin, après cette première expérience, notre mémoire de traduction s'est enrichie de nouveaux termes issus de l'application PhotoShop 3.0.5.

CONCLUSION

Les objectifs du projet ALAMET sont très ambitieux. Mettre en place une chaîne de traduction multi-plateformes est un projet de grande envergure. Aucun système de ce type n'est disponible, ce qui fait que ce projet s'aventure en terrain relativement vierge.

Dans cet article, nous avons seulement posé le problème, décrit les outils existants, et proposé une première solution. Bien que spécifique, elle nous a permis de nous rendre compte de tous les problèmes qu'on peut rencontrer, et a montré concrètement les avantages de notre méthode.

Le format TRADE et l'architecture du système proposé permettent, pour la première fois semble-t-il, de prendre en compte tous les problèmes posés par la localisation automatique.

Pour la suite, nous voulons intégrer dans la structure de la mémoire de traduction des concepts intéressants, comme les bases lexicales par acceptions, définies dans Sérasset (1994) et implémentées par Blanc (1995) dans PARAX.

RÉFÉRENCES

- Apple Computer Inc. (1996) : «*AppleGlot*», *User manual*, CD Tool Chest, Developer CD series, août 1996.
- ARCHIBALD, J. & A. DARISSE (1981) : *A Guide to Multilingual Publishing*, DC : Society for Technical Communication.
- BLANC, É. (1995) : «Une maquette de base lexicale multilingue à pivot lexical : PARAX», *Lexicomatique et dictionnaires*, Actes des IV^{es} Journée scientifiques du réseau LTT, Lyon, France, septembre 1995, Beyrouth, FMA et AUPELF-UREF, pp. 43-58.
- BULL, S.A. (1990) : «Lignes directrices pour le développement des produits internationaux», *Manuel de présentation du système ILO*, Bull S.A., CEDOC, février 1990.
- Microsoft Corp. (1996) : «*RL Tools*», *Users Manual*, <ftp.microsoft.com/pub/Dev/>.
- SEMMAR, N. (1995) : *Méthodologie de localisation des applications informatiques multimédia*, Nouvelle thèse, Université Paris VI, octobre 1995, 196 p.
- SÉRASSET, G. (1994) : *SUBLIM : un Système Universel de Base Lexicales Multilingues et NADIA : sa spécialisation aux bases lexicales interlingues par acceptions*, Nouvelle thèse, UJF, décembre 1994, 194 p.
- PLANAS, E. & Ch. BOITET (1997) : «Towards an Evolution of Memory Based Translation Systems», *PACLING'97*, Meisei University, Ohme, Tokyo, September 1997, 11 p.

QUAND LES MOTS PERDENT LA MÉMOIRE

Muriel CORET

Université Paris 7 - Denis Diderot, Paris, France

1. LA MÉMOIRE MORPHOLOGIQUE

On peut parler d'une mémoire des mots. Ou plutôt de mémoire de la langue qui conserve certaines informations concernant les mots. Je m'intéresserai ici à la mémoire qu'on pourrait appeler *morphologique*.

Comment rendre compte de l'abandon de *extradation*, recensé dans le *Vocabulaire des mots nouveaux* de Mercier (1801) au profit de *extradition*¹, forme attestée dans les dictionnaires du XIX^e siècle ? Si la première forme proposée peut sembler parfaitement régulière dans un premier temps (comparons *extrader*, *extradation* à *réparer*, *réparation*...) la seconde, elle, a l'avantage de garder en mémoire l'origine du mot par la présence du *-i-* de la série *trado*, *tradere*, *traditum*. Le choix de la forme *extradition* a donc l'avantage de rappeler l'origine du mot : ce choix est la manifestation de l'activité d'une mémoire.

Ce phénomène n'est pas un cas isolé. De nombreux aspects du fonctionnement du lexique révèlent que la langue conserve en mémoire certaines propriétés morphologiques du système latin, qu'elle peut réutiliser en français moderne pour de véritables créations. Il en est ainsi, par exemple, du phénomène de l'allongement thématique, observable dans le lexique du français et défini par Huot (1994) comme l'allongement d'une racine lexicale en *-(a/i)-t/s-* formant une base apte à certaines dérivations. C'est bien aux modèles d'alternances latines *amare* / *ama-t-um*, *mittere* / *miss-um*, *capere* / *cap-t-um*... que renvoient par exemple les alternances qu'on observe dans *aim-er* / *amat-eur*, *permett-re* / *permiss-ion*, *recev-oir* / *récept-ion*...

Mais comme toute mémoire, celle de la langue connaît des défaillances. Ce sont ces pertes de mémoire que je vous propose d'illustrer ici avec le cas des mots en *-ion*.

2. MÉMOIRE DES MOTS ET FORMATION DES NOMS EN *-ION*

La formation des mots en *-ion* constitue un bon exemple de cet espace de mémoire, en ce sens qu'elle est fondamentalement calquée sur la morphologie latine. On

¹ L'exemple est emprunté à M. Glatigny, «Néologisme et lexicographie», Forum de Morphologie, Lille, avril 1997.

observe une grande régularité en ce qui concerne la morphologie de ces noms : le suffixe *-ion* sélectionne des bases dites «thématiques», comportant dans la grande majorité des cas un allongement du radical en *-(a/i)t-*. Il en est ainsi dans *programm-at-ion*, *répét-it-ion*, *attribu-t-ion...*, qui comportent un allongement, mais aussi dans *production*, *concession*, *adhésion...* qui ne présentent pas en surface un véritable allongement mais dont les bases doivent être considérées néanmoins comme des bases thématiques², s'opposant aux formes non thématiques attestées dans les verbes correspondants *produi-re*, *concé-d-er*, *adhér-er...*

On retrouve là un trait de la morphologie du latin : la base qui apparaît dans le mot en *-ion* est directement calquée sur la formation latine des noms en *-io*, *ionis*, dans lesquels on retrouve le thème du supin. Exemples :

fr. N-ion	lat. N-io, ionis	Supin
<i>cession</i>	<i>cessio</i>	<i>cessum</i>
<i>gestion</i>	<i>gestio</i>	<i>gestum</i>
<i>correction</i>	<i>correctio</i>	<i>correctum</i>

Celle qui apparaît dans le verbe est issue, par dérivation savante ou populaire, du radical de l'infectum correspondant :

fr. V-er	V latin
<i>céder</i>	<i>cedere</i>
<i>gérer</i>	<i>gerere</i>
<i>corriger</i>	<i>corrigere</i>

97 % des mots en *-ion* du Robert électronique³ relèvent de ce fonctionnement : une même racine se présente sous une double forme, opposant une base thématique pour la suffixation en *-ion* (mais aussi en *-if*, *-oire*, *-ure...*) et une base non thématique pour le verbe correspondant (et la suffixation en *-age*, *-ment...*).

Or, on est amené à isoler un ensemble de mots qui résistent à cette analyse : ils comportent des bases qui, contrairement au cas général, ne présentent aucune variation. Ainsi, à la série régulière *programm-at-ion / programm-er*, *répét-it-ion / répét-er*, *attribu-t-ion / attribu-er* s'opposent des mots en *-ion* présentant le même radical que les verbes qui leur sont apparentés, tels que *abras-er / abras-ion*, *incis-er / incis-ion*, *invent-er / invent-ion...* au lieu de l'opposition «base thématique + *ion*» vs «base non thématique + *er*».

88 couples de mots, *N-ion* et *V-er*, sont dans ce cas. Ils sont regroupés, par famille, dans les 44 entrées du tableau donné ci-dessous en annexe. La totalité des couples «*V-er / N-ion*» de même base y figurent, exhaustivement, accompagnés d'un certain nombre d'informations commentées dans la suite du travail.

Ces mots posent problème. Que s'est-il passé ? La langue dispose-t-elle de deux manières de construire les noms en *-ion* ? Faut-il considérer que ces mots en *-ion* relèvent d'une autre règle que celle décrite plus haut (et donc en restreindre la généralité) ?

2 Ce point ne sera pas développé ici. Cf. Coret (1994).

3 Le *Grand Robert électronique* comporte 2951 mots en *-ion*.

3. RECONSTRUCTION DE VERBES ET PERTE DE MÉMOIRE

Mon analyse s'appuie sur des arguments d'ordre historique, sémantique et morphologique. Elle a permis de mettre à jour les caractéristiques suivantes :

Aspect historique

Une vérification rapide montre que les noms en *-ion* sont apparus relativement récemment dans les dictionnaires (et donc, nous l'admettons, dans la langue) mais sont généralement donnés comme antérieurs aux verbes correspondants. Un tiers environ des mots considérés ne figurent pas du tout dans le dictionnaire Richelet. On peut mentionner à titre d'exemple :

	Datation ⁴ du <i>N-ion</i>	Datation du <i>V-er</i>
<i>collect</i>	XIV ^e	XVI ^e
<i>concoct</i>	1528	XX ^e
<i>digress</i>	XII ^e	1838
<i>édit</i>	XIII ^e	1878
<i>excrét</i>	XVI ^e	1836
<i>inject</i>	XIII ^e	XVIII ^e
<i>inspect</i>	1290	1781
<i>lés</i>	XII ^e	1611
<i>réfract</i>	XVI ^e	1752
<i>révuls</i>	XVI ^e	1845
<i>transit</i>	XIII ^e	1832
<i>vis</i>	XII ^e	1668

Dans ce domaine bien sûr, une uniformité totale n'est pas possible. Mais, comme le souligne déjà Huot (1997), l'important est que, majoritairement, ces *N-ion* ne peuvent pas être considérés comme issus des verbes en *-er* et que, d'autre part, la formation des *V-er* est encore un processus vivant aujourd'hui (comme l'attestent quelques formations récemment apparues dans les dictionnaires comme *compacteur*, *collecter*...).

Aspect sémantique

On constate que le sens des verbes formés sur la base des noms en *-ion* est décrit de manière très générale par les dictionnaires à l'aide de la paraphrase⁵ «faire / commettre / procéder à [l'action exprimée par le *N-ion*]» — ce qui n'est pas le cas pour les autres couples *V-er* et *N-ion*, pour lesquels au contraire, le verbe sert à définir le nom (paraphrases en «action de V»). On relève par exemple⁶ dans le *Petit Robert* :

4 Les datations sont celles de Littré (1863) ou du dictionnaire *Le Petit Robert*.

5 Les dictionnaires montrent sur ce point une telle hétérogénéité que la moindre régularité mérite toute notre attention !

6 Il est à noter que lorsque la définition ne mentionne pas le *N-ion*, elle a généralement recours à une autre unité comportant la base thématique. Cf. *collecter* : «faire une collecte»; *compacteur* : «rendre compact»; *exempter* : «rendre exempt»; *professer* : «enseigner en qualité de professeur»; *propulsion* : «faire avancer à l'aide d'un propulseur»...

<i>abraser</i> :	«User (une matière, un objet) par abrasion »
<i>agresser</i> :	«Commettre une agression »
<i>confesser</i> :	«Entendre (un fidèle) en confession »
<i>connecter</i> :	«Unir par une connexion »
<i>décompresser</i> :	«Cesser ou diminuer la compression de»
<i>déconnecter</i> :	«Supprimer la connexion de»
<i>décontracter</i> :	«Faire cesser la contraction musculaire de»
<i>diffracter</i> :	«Produire la diffraction de»
<i>digresser</i> :	«Faire des digressions »
<i>exciser</i> :	«Enlever par excision »
<i>excréter</i> :	«Évacuer par excrétion »
<i>exploser</i> :	«Faire explosion »
<i>imploser</i> :	«Faire implosion »
<i>impulser</i> :	«Donner une impulsion à»
<i>objecter</i> :	«Opposer (une objection) à»
<i>opiner</i> :	«Dire, énoncer son opinion »
<i>perfuser</i> :	«Pratiquer une perfusion sur»
<i>préciser</i> :	«Apporter des précisions »
<i>presser</i> :	«Exercer une pression »
<i>rééditer</i> :	«Donner une nouvelle édition de»
<i>réfracter</i> :	«Faire dévier par le phénomène de la réfraction »
<i>régresser</i> :	«Subir une régression »
<i>révulser</i> :	«Faire affluer par révulsion le sang de»
<i>sécréter</i> :	«Produire (une substance) par sécrétion »
<i>transfuser</i> :	«Soumettre à une transfusion sanguine»

Ces définitions mettent en évidence une valeur particulière pour ces verbes qu'on pourrait caractériser de verbes «d'effectuation». Ce fait, déjà souligné dans Coret (1994) et Huot (1994), mériterait d'être étudié dans le cadre d'une hypothèse plus générale sur la valeur du segment d'allongement lui-même, qui *n'exprimerait rien d'autre*, selon Huot (1994 : 55) *que l'accompli*.

Aspect morphologique

Ces noms, s'ils peuvent paraître dans un premier temps mal formés en français, sont parfaitement réguliers en regard du modèle latin. Ils présentent une base thématique, issue du thème du supin latin qui, dans presque tous les cas⁷, alternait avec une forme simple, conformément aux règles morphologiques du système. Cette alternance figure dans la troisième colonne du tableau en annexe. Citons par exemple :

Français	Supin latin	Infinitif latin
<i>ablat-</i>	<i>ablatum</i>	<i>aufferre</i>
<i>abras-</i>	<i>abrasum</i>	<i>abradere</i>
<i>accept-</i>	<i>acceptum</i>	<i>accipere</i>
<i>agress-</i>	<i>aggressum</i>	<i>aggredi</i>

7 Ce n'est pas le cas en effet pour *annexio*, *communio*, *complexio*, *opinio*, *rebellio*, *recensio*, qui n'ont qu'une base. Ce sont là les seuls cas où on ne puisse pas reconstruire une base non thématique.

On peut voir qu'il s'agit de formations régulières à l'origine, et on retrouve bien ici l'application de la règle générale citée plus haut : le *N-ion* est formé sur une base thématique, qui alterne avec une base simple. La particularité que je veux pointer ici, c'est que ces bases n'ont subsisté que sous leur forme thématique.

Pour autant, la base non thématique (de formation populaire ou savante) n'est pas totalement éliminée : dans 38 cas sur 44, une base non thématique était accessible pour la formation du verbe et rien, du point de vue morphologique, n'empêchait sa réactivation. Le lexique présente en fait différentes traces de son existence :

- il arrive qu'on retrouve un verbe formé sur la base non thématique dans un état de langue ancien. Ainsi *affaiter*, *colliger*, *expeller*, *invertir*, *objicer*, *transir* (face à *affect-*, *collect-*, *expuls-*, *invers-*, *object-*, *transit-*) ont été attestés en français et sont mentionnés dans des dictionnaires anciens ou des dictionnaires historiques.
- la base non thématique a pu aussi subsister de manière isolée sous la forme d'un nom ou d'un adjectif encore attesté en français moderne : *diffract-* / ***diffring-****ent*, *divis-* / ***divid-****ende*, *incis-* / ***incid-****ence*...
- la base non thématique a pu enfin se maintenir en français moderne, sous la forme d'un premier verbe (normalement constitué), qui se trouve donc en concurrence avec le *V-er* sur base thématique récemment formé. C'est ainsi qu'on a les concurrents
cueillir (de *colliger*) / *collecter*
comprimer / *compresser*
dire / *dicter*
repousser / *répulser*
voir / *viser*...

L'existence d'un premier verbe (construit sur base non thématique) n'a pas bloqué la création d'un second verbe. Ils sont issus de la même racine et on pourrait considérer que, à un certain niveau d'analyse, ils ont le «même» sens. Pourtant, leur coexistence même empêche de les considérer comme des synonymes. En fait, une fois le premier verbe entré dans le lexique, il acquiert un sémantisme particulier, qui l'éloigne de la simple relation au *N-ion*. C'est cette rupture que pallie la création d'un nouveau verbe, sur la base exacte du *N-ion*.

Dans les cas où on ne peut pas retrouver cette base non thématique dans un mot attesté, il est possible de reconstruire une base possible (que nous noterons par le symbole °) d'après le modèle général d'évolution de la langue, en prenant éventuellement appui sur d'autres séries de mots, qui ont, elles, conservé les deux termes de l'alternance. On pouvait ainsi attendre °*concuire* (d'après *cuire*), °*concroître* (*croître*), °*contraire* ou °*contirer* (*traction* / *traire*, *tirer*), °*détéger* (*protection* / *protéger*), °*diffondre* (*fusion* / *fondre*), °*excider* (*décision* / *décision*), °*invenir* (*prévention* / *prévenir*)...

Seuls se présentent en fait comme cas particuliers les bases *annex-*, *commun-*, *complex-*, *opin-*, *rebell-*, *recens-*, pour lesquelles on ne peut pas reconstruire une base non thématique : elle n'est attestée dans aucun mot du français et ne peut pas être reconstruite d'après le supin puisque en latin déjà, la même base était attestée pour la formation du supin et de l'infinitif correspondant.

4. MÉMOIRE MORPHOLOGIQUE ET RÉGULARITÉ LEXICALE

Ces observations montrent qu'on peut rendre compte de cet ensemble de cas sans remettre en cause la règle générale et sans avoir recours à la notion d'exception. Cela passe par la reconnaissance d'une perte de mémoire de la langue, qui n'est pas sémantique — ce qui est un phénomène déjà bien établi — mais morphologique.

Il s'opère une «reconstruction» : contrairement à l'ensemble des autres mots en *-ion*, ces 88 mots ne peuvent pas être analysés comme dérivés d'une base verbale «préexistante». Ce sont au contraire les verbes qui sont reconstruits à partir des mots en *-ion* : sur la base thématique, extraite d'un mot en *-ion* régulièrement formé (base thématique), la langue reconstruit un verbe du premier groupe. Une fois entrée dans le lexique, la base empruntée s'impose pour tout le paradigme et peut donner lieu à la construction d'un verbe (*collecter*), d'un nom (*collectage*), etc.

Arrêtons-nous sur la série *collectage*, *collecter*, *collection*, particulièrement représentative de la situation. Cette série est en fait le résultat de découpages successifs dans le temps, qui aboutissent à une superposition de formations qui, du strict point de vue synchronique, pourrait sembler déroutante. Partant d'une base thématique *collec-t-*, empruntée au latin (supin *collectum*), la langue produit normalement *collec-t-ion*, sur base thématique et *collig-er*, sur base non thématique. Mais il arrive qu'une base cesse de fonctionner en tant que telle et ne soit plus reconnue. Ainsi, l'alternance *collec-t- / collig-* n'est plus perçue et fait place à la seule base *collec-t-*, disponible pour former de nouveaux mots : *collect-er* (signalé comme fautif dans le Larousse du XX^e siècle) et *collect-age* (d'apparition récente).

On retrouve bien ici la situation observée pour l'ensemble des couples *N-ion* et *V-er* sur bases thématiques et le jeu complexe de la mémoire de la langue⁸.

On peut donc parler d'une perte de mémoire, dans la mesure où, lors de ces reconstructions, la langue «oublie» deux choses :

- l'existence, le cas échéant, d'une base non thématique (ou la possibilité d'en déduire une par comparaison avec d'autres familles de mots) et la possibilité de l'utiliser comme une base simple du français;
- le caractère thématique de la base utilisée : une fois le verbe créé, la base semble perdre son caractère thématique et se trouve alors disponible pour une dérivation avec un suffixe tel que *-age* (*collectage*, *détritage*), qui sélectionne des bases non thématiques.

Les observations présentées ici ont pour conséquence que ces mots en *-ion* ne doivent pas être considérés comme exceptionnels : la règle générale qui décrit le processus de suffixation en *-ion* n'est pas remise en cause.

8 La même analyse rend compte de tous les *N-age* comportant la même base qu'un *N-ion* : *compactage*, *détection*, *détritage*, *factage*, *fruitage*, *télédictage*, *visage*. Cf. Coret (1994), ainsi que d'autres verbes construits sur base thématique.

Mais au-delà de la simple explication du sous-ensemble de mots considéré (dont il faut bien de toutes façons rendre compte) cette analyse a mis en lumière un aspect particulier de la création lexicale — l'extraction de bases — et illustre aussi, en même temps que l'existence d'une mémoire des mots, son caractère lacunaire.

RÉFÉRENCES

CORET, M. (1994) : *Problèmes de suffixation et structuration du lexique*, Thèse de doctorat, Université Paris 7.

Dictionnaire Le Robert (1996) : version CD-ROM, Paris, Le Robert.

Le Grand Robert électronique (1989) : version CD-ROM, Paris, Le Robert.

HATZFELD, A., DARMESTETER, A. et A. THOMAS (1890-1900) : *Dictionnaire général de la langue française*, Paris, Delagrave.

HUOT, H. (1994) : «Sur la notion de racine», *TAL* 35-2.

HUOT, H. (1997) : «Des mots possibles aux mots existants : système morphologique et structuration du lexique», *Sillexicales 1*, Lille, Université de Lille 3.

LITTRÉ, É. (1863) : *Dictionnaire de la langue française*, Paris, Gallimard - Hachette.

MERCIER (1801) : *Néologie ou Vocabulaire des mots nouveaux*.

PICOCHÉ, J. (1989) : *Dictionnaire étymologique du français*, Paris, Le Robert.

REY, A. (1992) : *Dictionnaire historique de la langue française*, Paris, Le Robert.

RICHELET, F. (1693) : *Dictionnaire françois contenant généralement tous les mots*, Genève.

ANNEXE : BASES THÉMATIQUES COMMUNES À *N-ION* ET *V-ER*

Le tableau ci-dessous regroupe la totalité des bases du français figurant à la fois, sans variation, dans un *N-ion* et un *V-er*. Pour chacune d'elle, j'indique l'étymologie latine (telle qu'elle est indiquée par Littré ou le *Petit Robert*), qui révèle bien, dans la grande majorité des cas, une opposition entre une base thématique et une base non thématique.

Dans les quatrième et cinquième colonnes figurent la base non thématique reconstruite ou attestée, disponible en théorie, et des mots révélateurs de la survivance de cette base non thématique. Les verbes possibles non attestés sont notés à l'aide du symbole °; ceux attestés en ancien français sont précédés de la mention «AF». La base non thématique pouvant résulter d'une formation populaire ou savante, on mentionne quelquefois deux mots. C'est le cas par exemple pour *collect-*, face à qui on a *collig-er* (formation savante) et *cueill-ir* (formation populaire).

	Base de <i>N-ion / V-er</i>	Etymologie latine du <i>N-ion</i>	Base non thématique	Mots attestés et verbes possibles (sur base non thématique)
1	<i>ablat</i>	ablatio, au-ferre	-fer-	cf. <i>conférer, déférer, souffrir</i> . issus de V latins en -ferre
2	<i>abras</i>	abrasio, ab radere	-rad- / -rat-	° <i>abrader</i> (cf. <i>radoire</i> , AF <i>ratoire</i>)
3	<i>accept</i> <i>except</i> <i>intercept</i>	acceptio, accipere exceptio, excipere interceptio, interciperere, capere	-cip- / -cev-	° <i>accevoir</i> (cf. <i>déception / décevoir</i>)
4	<i>affect</i> <i>infect</i> <i>désinfect</i> <i>réinfect</i>	affectio, afficere, facere infectio, infectus, inficere, facere mot préfixé, cf. infect- mot préfixé, cf. infect-	-fai-	AF <i>affauter</i> : ° <i>affaire</i> (cf. <i>perfection / parfaire</i>)
5	<i>agress</i> <i>digress</i> <i>progress</i> <i>régress</i> <i>transgress</i>	adgressio, adgredi, gradi digressio, digressum, digredi, gradi progressio, progredi, gradi regressio, regressus, regredi, gradi transgressio, gradi	-grad-	cf. <i>rétrograder, dégrader</i> issus de verbes latins en -gradi
6	<i>annex</i> <i>désannex</i> <i>connect/x</i> <i>déconnect/x</i>	annexio, annectere mot préfixé, cf. annex- connexio, connectere mot préfixé, cf. connect-		une seule base en latin
7	<i>assert</i> <i>désert</i>	assertio, asserere desertio, deserere	-ser-	° <i>asserer</i> (cf. <i>insertion / insérer</i>)
8	<i>collect</i>	collectio, colligere	-lig-	<i>colliger</i> ; <i>cueillir</i>
9	<i>commun(i)</i>	communio, communis		une seule base en latin
10	<i>compact</i>	compactus, compingere, pangere	-ping-	° <i>compinger</i> ?
11	<i>complét</i>	completus, complere, plere	-pli-	° <i>complir</i> (cf. <i>emplir, remplir</i> issus du latin -plere)
12	<i>complex</i>	complexio, complexus, complexere		une seule base en latin
13	<i>compuls</i> <i>expuls</i> <i>impuls</i>	compulsio, compulsus, compellere expulsio, expellere impulsio, impulsus, impellere, pellere	-pell- / -pouss-	AF <i>expeller, pousser</i> (cf. <i>pousser</i> du latin -pulsus)

Quand les mots perdent la mémoire

	<i>propuls</i>	pro pulsum, pellere		
	<i>répuls</i>	repulsio, repulsum, repellere		
14	<i>concoct</i>	concoctio, coquere	-cu(s)-	°concuire (cf. cuire, décuire)
15	<i>concrét</i>	concretio, concrecere	-croît-	°concroître (cf. croître)
16	<i>confess</i>	confessio, confiteri	-fir-	cf. confiteor
	<i>profess</i>	professio, professus, profiteri		
17	<i>contract</i>	contractio, contrahere, trahere	-traï- / -tir-	°contraire / °contrer (cf. traction / trave ; trer)
	<i>décontract</i>	mot préfixé, cf. contract-		
	<i>détract</i>	detractio, trahere		
18	<i>délect</i>	detectum, detegere	-rég-	°détéger (cf. protection / protéger)
19	<i>détrit</i>	detratio, deterere	-ter-	°déterer ?
20	<i>dict</i>	dictio, dicere	-di-	dire
	<i>édict</i>	edictum		
21	<i>diffract</i>	diffRACTus, diffringere	-fring-	°diffringer (cf. diffringent, réfringent)
	<i>réfract</i>	refractio, refringere		
22	<i>diffus</i>	diffusum, diffundere	-fond-	°infondre (cf. fusion / fondre)
	<i>infus</i>	infusio, infusus, infundere		
	<i>perfus</i>	perfusio, perfusum, perfundere		
	<i>transfus</i>	transfusio, transfundere		
23	<i>dispers</i>	dispersio, dispergere, spargere	-perg-	°disperger (cf. asperston / asperger)
24	<i>divis</i>	divisio, dividere	-vid-	°divider (cf. dividende)
	<i>subdivis</i>	mot préfixé, cf. divis-		
25	<i>édit</i>	editio, edere	-di-	°édre (cf. dire, dédire)
	<i>réédit</i>	mot préfixé, cf. édit-		
26	<i>éject</i>	ejectio, ejectum, ejicere, jacere	-jic- / jet-	AF objicer, objeter , interjeter (cf. jeter)
	<i>inject</i>	injectio, injicere		
	<i>interject</i>	interjectio, interjicere		
	<i>introject</i>	emprunté à l'allemand		
	<i>object</i>	objectio, objectum, objicere		
	<i>réinject</i>	mot préfixé, cf. inject-		

27	<i>électrocute</i> <i>exécute</i> <i>persécute</i>	emprunté à l'anglais exsecutio, exsequi persecutio, per sequi	-séqu- / -suiv-	cf <i>consécution / conséquence ; suivre</i>
28	<i>excise</i> <i>incise</i> <i>précise</i>	excisio, excisum, excidere, caedere incisio, incisum, incidere praecisio, praecisus, praecidere, caedere	-cid-	^o <i>excider</i> (cf. <i>décision / décider ; incident</i>)
29	<i>excrète</i> <i>sécrète</i>	excretio, excretum, excernere, cernere secretio, secernere, cernere	-cern-	cf <i>décerner, concerner</i> .. issus du latin -cernere
30	<i>exempt</i>	exemptio, exemptum, eximere	-im-	^o <i>eximer</i> (cf <i>péremption / périmer, rédimmer</i>)
31	<i>explos</i> <i>implos</i>	explosio, explosum, explodere, plaudere formé sur explosion	-plod-	^o <i>exploder</i> (cf <i>applaudir</i> du latin applaudere)
32	<i>inspect</i> <i>prospect</i>	inspectio, inspicere prospectus, pro specere	-spic-	cf <i>perspicace, auspice</i>
33	<i>invent</i>	inventio, inventum, invenire	-ven-	^o <i>invenir</i> (cf <i>prévention / prévenir</i>)
34	<i>invers</i>	inversio, invertere	-vert-	<i>invertir</i>
35	<i>lés</i>	laesio, laesus, laedere	-lid-	cf <i>élision / éluder</i>
36	<i>opin</i>	opinio, opinari		une seule base en latin
37	<i>opt</i> <i>adopt</i>	optio, optare, optare adoptio	-op-	^o <i>opter</i> ?
38	<i>préfix</i>	praefixus, praefigere	-fig- / -fich-	cf <i>ficher</i> du latin figere
39	<i>press</i> <i>compress</i> <i>décompress</i>	pressio, primere compressio, compressum, comprimere mot préfixé, cf compress-	-prim-	<i>comprimer</i>
40	<i>rébell</i>	rebellio, rebellis		une seule base en latin
41	<i>recens</i>	recensio, recensere		une seule base en latin
42	<i>révuls</i>	revulsio, revulsum, revellere, vellere	-vell-	^o <i>réveller</i> ?
43	<i>transit</i>	transitio, transire	-ir-	<i>transir</i>
44	<i>vis</i> <i>révis</i> <i>supervis</i>	visio, visum, videre visum, revisere mot préfixé, cf. vis-	-vid- / -voi-	<i>voir</i>

L'IMAGE ET LA FORME : APLATISSEMENT OU DISTORSION DU TEMPS ?

Xavier LELUBRE

CRTT, Université Lumière Lyon 2, Lyon, France

1. LA TRACE DES ÉTATS ANTÉRIEURS DES CONNAISSANCES

L'image que nous avons d'une science et de sa terminologie, au moment présent qui est le nôtre, n'est-elle pas la forme figée qu'elle a alors prise ?

1.1 Chez Ibn al-Haytam, l'image et la forme

Le point de départ de cette recherche est l'existence de deux termes arabes d'optique employés de nos jours concurremment, qui correspondent au terme français (et aussi anglais) *image* : il s'agit de *sûra*¹ (c'est le terme le plus généralement employé dans le monde arabe, à commencer par l'Égypte) et *hayâl* (employé surtout en Syrie)².

Ce qui attire notre attention ici est le fait que ces deux termes, loin d'être d'utilisation récente en optique, se trouvaient déjà utilisés au Moyen Âge, mais où ils avaient deux acceptions différentes.

Dans la langue générale, *hayâl* a le sens de «fantôme, spectre, ce qui apparaît à un homme éveillé ou en songe»... et c'est aussi l'image que l'on voit dans un miroir³. Quant à *sûra*, c'est la forme (syn. *sakl*); extérieur, aspect, apparence; manière, façon (syn. *wazh*); figure, image (représenté par la peinture ou le dessin).

¹ La transcription des caractères arabes utilisée ici est, dans l'ordre de l'alphabet oriental arabe habituel : / ? b t ṭ z h ḥ d ḍ r z s ṣ d t z ` g f q k l m n h w y l.

² L'Égypte et la Syrie ont été lors de la Renaissance arabe au siècle dernier (la *Nahda*) les grands pourvoyeurs de terminologie scientifique.

³ Le grand dictionnaire *Lisân al-`Arab* d'Ibn Manzûr (m. 1311), donne, entre autres : «*al-hayâl^u li-kullⁱ ṣay[?]in tarâ-hu ka z-zillⁱ, wa-kadâlika hayâl^u l-?insânⁱ fî l-mir?âtⁱ wa-hayâl^u-hu fî l-manâmⁱ ṣarat^u timtâlⁱ-hi* » [ce qui apparaît de toute chose que l'on voit, comme l'ombre, de même, l'image de l'homme dans le miroir, et aussi : son spectre dans le sommeil est la forme sous laquelle il est représenté].

Nous pouvons voir la différence entre ces deux termes chez le célèbre savant arabe Ibn al-Hayṭam (m. vers 1039)⁴, qui a joué au Moyen Âge un rôle fondamental dans l'évolution de l'optique, dans le passage suivant (cité par Mustafâ Nazîf 1942-1943 : 596)⁵ :

«L'image (*ḥayâl*) est la forme (*sûra*) de la chose vue (*mubsar*) que le regard (*basar*) saisit par la réflexion sur la surface [de séparation] du corps poli. Le lieu de cette image est le lieu où le regard saisit cette forme».

L'acception prise par ces lexies en optique ne peut être appréhendée que dans le cadre épistémologique où elles ont été établies comme termes. Le savant arabe établit une théorie de la lumière, une théorie de la vision⁶, se démarquant dans son ouvrage, en particulier de la théorie antérieure des «rayons visuels», soutenue par différentes écoles depuis l'Antiquité.

1.2 Évolution des connaissances et terminologie

Comme le note Alain Rey (1979 : 64), avec l'évolution constante de leurs configurations conceptuelles, les sciences voient «leur terminologie évoluer, mais conserver forcément la trace des états antérieurs des connaissances : c'est donc un rapport changeant entre termes en partie anciens et notions nouvelles que la terminologie scientifique doit définir. Il s'agit ici avant tout d'une mise au point permanente des définitions.»

Ce sera bien le cas en optique, où des concepts physiques fondamentaux, comme ceux de lumière, de rayon lumineux, etc. ne cesseront d'être appréhendés, retravaillés, réinterprétés dans le cadre de théories concurrentes⁷, tout en gardant les mêmes dénominations.

Ces dénominations pourraient-elles conserver une trace, en quelque sorte la mémoire de leurs acceptions antérieures ?

La mémoire est bien sûr liée au temps. Le déroulement du temps dans les civilisations n'est ni continu ni homogène. Il n'est qu'un lointain effet du temps physique exprimé par la succession des jours et des saisons et du temps exprimé par la succession des générations. C'est à cela que fait allusion le titre de cet article.

⁴ Connu en Occident sous le nom d'Alhazen, son ouvrage principal, *Kitâb al-Manâzîr* (Le livre de l'optique) sera traduit en latin (par Risner, en 1572), sous le nom d'*Opticae Thesaurus Alhazeni Arabis*, ouvrage qui eu une influence considérable dans ce domaine en Europe.

⁵ Mustafâ Nazîf (1942-1943 : 103) indique comment, chez cet auteur, *sûra* a le sens du terme chez les philosophes (la *forme*, opposée à la *substance* (*hayûlâ*) et *ḥayâl*, c'est l'*image* dans l'acception moderne. Voir aussi, sur *sûra* (chez Ibn al-Hayṭam) traduit par *forme*, Roshdi Rashed (1970 : 278-280).

⁶ Sur l'optique chez Ibn al-Hayṭam, voir en particulier Gérard Simon (1989), Roshdi Rashed (1993).

⁷ Ibn al-Hayṭam développe une théorie de la lumière. Bien plus tard viendront les théories électromagnétiques, corpusculaires, quantiques sur la nature de la lumière, toujours dénommée, au-delà de ces approches, *lumière*. D'un autre point de vue, sur les termes latins, de *lumen* et *lux*, pour la lumière, voir Vasco Ronchi, 1966 : 103.

Le temps ici considéré est sa matérialisation dans l'évolution des composantes d'une civilisation donnée, et pour ce qui nous intéresse plus particulièrement, parmi les domaines de la connaissance scientifique et de la pratique technologique, celui de l'optique.

L'évolution de ces différents domaines est généralement discontinue et différentielle, chacun évoluant à sa vitesse, avec des pauses, des accélérations, voire au contraire des retours en arrière. Il y a forcément distorsion entre le *temps* d'un domaine donné et le *temps* de la société.

Cette évolution est notamment liée à celle des autres domaines — ainsi l'optique et les mathématiques, l'astronomie, la fabrication des composants optiques, le développement d'autres branches de la physique, comme la mécanique, l'électromagnétique...—, à l'état général de la société considérée, et à ses relations avec d'autres civilisations, directement ou indirectement — par les traductions — l'optique en Occident s'est développée au Moyen Âge en grande partie sur la base des traductions latines de documents en arabe, eux-mêmes redevables aux traductions en syriaque puis en arabe de documents grecs de l'Antiquité⁸.

Les connaissances, les théorisations, les concepts, les savoir-faire, les procédés, les outils constitutifs de ces domaines — les *notions* ou *concepts* de la terminologie, que nous appelons *unités référentielles* — s'expriment par les terminologies qui leur sont associées.

Les unités référentielles d'un domaine sont liées à un état donné de la constitution de ce domaine. Leur histoire relève de l'histoire des civilisations, des idées, des sciences et techniques. Quant aux terminologies qui leur sont associées, elles constituent le matériau de cette étude.

2. L'OPÉRATION DE DÉNOMINATION ET LA CAPTATION D'EFFETS DE MÉMOIRE

L'unité référentielle d'un domaine donné, établie par ceux qui l'ont «découverte», l'est avec sa dénomination. Celle-ci peut se trouver modifiée par d'autres instances — et elle l'est bien évidemment quand elle a à être dénommée dans d'autres langues que celle de ceux qui ont établi l'unité référentielle.

La question se pose de savoir si lors de l'opération de dénomination d'une unité référentielle, le mécanisme lui-même de cette opération ne conduirait pas à capter des éléments extra-terminologiques, qui correspondraient à une époque et à un moment de l'évolution de la spécialité concernée, trace d'événements qui y sont reliés.

Ces traces pourraient constituer la *mémoire* d'un terme.

La mémoire des termes serait-elle alors l'un des constituants de la partie inavouable, soigneusement évacuée par la terminologie, leur connotation ? C'est-à-dire le lien d'un terme avec les autres mots de la langue, au-delà des rapports de ce terme avec les autres termes du «système terminologique» — appellation d'ailleurs impropre⁹ — dont il ressort ?

⁸ Parmi une abondante bibliographie sur l'optique dans le Monde islamique, Bernard Maitte (1981 : 21-26), A. I. Sabra, «Manâzir ou `Ilm al-Manâzir», in *Encyclopédie de l'Islam*, tome IV (2^e éd.).

⁹ Ce qui fait système, ce sont les unités référentielles du domaine ou du sous-domaine référentiel, et non pas, généralement, les dénominations. Ainsi, en optique géométrique, ce

2.1 Le cas de la terminologie arabe

Si nous considérons l'état actuel de la terminologie arabe de l'optique, celle-ci a un fonds ancien auquel nous avons fait allusion précédemment. Mais le plus gros de cette terminologie est bien plus récent. En particulier, on sait, en ce qui concerne les terminologies scientifiques et techniques arabes contemporaines l'importance, pour des raisons historiques, du français et de l'anglais. C'est essentiellement à travers ces deux langues — et il faut tenir compte de ce que chacune est ou a été la langue de référence utilisée dans telle ou telle région du Monde arabe bien souvent à l'exclusion de l'autre —, que le Monde arabe contemporain participe au mouvement scientifique mondial.

Le premier contact avec les terminologies scientifiques arabes peut déconcerter le traducteur, tant est importante la question des phénomènes de variation terminologique d'une région arabe à une autre, voire d'un auteur à un autre.

La profusion de termes synonymes — que les instances interarabes s'attachent à limiter — est par elle-même signifiante. Elle ressort d'au moins trois faits :

- l'émiettement de la création terminologique, qui peut être considéré comme une conséquence du manque d'unité politique;
- l'influence de deux centres arabes créateurs de terminologie, l'Égypte et la Syrie;
- l'existence de deux sources étrangères de terminologie, la source française et la source anglaise.

C'est ce que nous disent, comme nous allons le voir, de nombreux termes de l'optique, tant par l'existence de séries de termes synonymes, que, par exemple, de faits de polyvalence ou d'hyponymie.

2.2 L'impondérable part de mémoire captée par les différents types de dénomination

L'opération de dénomination d'une unité référentielle consiste, si le terme est motivé, à exprimer un sous-ensemble¹⁰ des traits de substance, des caractéristiques de cette unité référentielle. Ce choix n'est pas livré au hasard.

Deux situations sont possibles : celle où la dénomination est faite dans la langue de celui ou ceux qui ont établi cette unité référentielle et celle où la dénomination est reprise dans d'autres langues¹¹. Dans ce dernier cas, deux possibilités existent : exprimer cette dénomination dans la langue d'accueil — au moyen des différents procédés possibles offerts par cette langue — ou bien procéder à une dénomination autonome, à partir de l'unité référentielle elle-même, par choix éventuellement d'autres traits de substances. Dans

sont l'objet, l'image, l'espace objet, l'espace image, le système optique, etc., qui font système, et non pas les termes *objet*, *image*,... Les procédés les plus divers de création de termes, la diachronie dans laquelle cette création s'inscrit; s'opposent à ce qu'ils constituent un système, sauf pour des sous-ensembles bien limités.

¹⁰ La dénomination est une opération extrêmement réductrice. Elle ne peut exprimer que quelques traits de substances de l'unité référentielle.

¹¹ Ce sont respectivement la *néonymie d'origine* et la *néonymie d'appoint* de Guy Rondeau (1983 : 124)

le premier cas, l'on s'appuie avant tout sur la dénomination de la langue source; auquel cas, il y a forcément, d'une façon ou d'une autre, une opération de calque : la dénomination est guidée par l'original.

En particulier, le cas du calque, qui peut fonctionner dans les deux sens. Ainsi au Moyen Âge, de l'arabe vers le latin, puis le français : la *ré tine* : *ṣabakiyya*, l'*humeur cristalline* : *rutûba zalîdiyya*, la *camera obscura* (la chambre obscure, aujourd'hui encore, c'est le terme latin, du Moyen Âge, qui est généralement utilisé en français), *bayt muzlim*.

L'influence de la langue d'origine peut s'exprimer aussi de bien d'autres manières, au niveau de la synonymie, à celui de la polyvalence ou de l'hyponymie.

Ainsi, pour la dénomination d'instruments à vision éloignée, le français appelle généralement *télescope* l'instrument dont l'objectif est un système réflecteur et *lunette* l'instrument dont l'objectif est un système réfracteur¹². Les équivalents anglais courants sont, respectivement, *reflecting telescope* et *refracting telescope*. L'anglais possède de ce fait un terme hyperonyme, *telescope*, tandis que le français... n'en possède pas¹³. Ces divergences terminologiques se réfractent, si l'on peut dire, dans les terminologies arabes, selon qu'elles suivent le français ou l'anglais.

À nouveau deux cas se présentent : ou la dénomination choisie est une lexie qui existe déjà dans la langue, mais à laquelle on donne une acception nouvelle, pour le domaine de spécialité considéré, ou bien c'est une lexie nouvelle qui est forgée. Dans ce dernier cas, elle utilisera des éléments de la langue — morphèmes ou lexies constituantes — existant déjà. Dans les deux cas, de toute façon, la nouvelle unité terminologique ne sera jamais sans lien avec d'autres. Bien sûr, ces liens diffèrent d'une langue à une autre.

La langue arabe dispose de quatre procédés pour créer ses termes : le recours à son système de nomination¹⁴ (les procédés morphologiques), le recours à son système de communication (les procédés syntaxiques), le recours aux transferts sémantiques et, enfin, l'emprunt.

a) le recours au système de nomination

Il aboutit à la formation d'unités terminologiques simples. Nous avons vu le cas de *ḥayâl*, *image*. La lexie est ancienne, attestée par la poésie, elle a plusieurs emplois et, en particulier, celui d'image vue dans un miroir. L'arabe ancien de spécialité a tout naturellement repris cette lexie, en en précisant l'acception dans le domaine de l'optique. Il en est ainsi pour de très nombreux termes.

¹² Les télescopes s'appellent aussi *réflecteurs* ou encore *télescopes catoptriques*. Les lunettes portent aussi le nom de *réfracteurs* ou *télescopes dioptriques* (H. Pariselle, *Les instruments d'optique*, Paris, Armand Collin, 1933 : 99)

¹³ On utilise alors l'expression «les lunettes et les télescopes».

¹⁴ Sur les systèmes de nomination et de communication de l'arabe, voir André Roman (1989).

En revanche, le terme *istiqtâb*, l'équivalent arabe de *polarisation*¹⁵, n'est pas attesté dans les dictionnaires arabes anciens. Il a été créé très probablement en référence à la lexie *qutb*, dont l'un des sens est *pôle*, de racine (*q - t - b*), comme nom d'action du verbe dérivé *istaqtaba*. Cette racine, elle, est ancienne et a plusieurs sens, qui nous importent peu ici. Mais on ne peut évacuer, pour le locuteur arabophone, les résonances de sens que peut induire la racine, ici triconsonantique, mais qui peut être quadriconsonantique¹⁶.

Il existe ainsi pour chaque terme arabe formé dans le cadre du système de nomination, c'est-à-dire, formé sur une racine et un schème, des relations inévitables avec les autres lexies ayant la même racine, ou voire le même schème¹⁷.

b) le recours au système de communication

Ce procédé aboutit à la formation d'unités terminologiques complexes, dont les éléments sont ou ne sont pas des termes.

Ainsi, à titre d'exemples, les termes *istiqtâb dâ ?iriyy yamîniyy*, *polarisation circulaire droite*, *mizhar tabâyun at-tawr*, *microscope à contraste de phase*.

Ce mode de formation est fréquemment sollicité en arabe. Celui-ci ne dispose pas, pour la formation de sa terminologie scientifique, d'une langue de prestige, comme le grec et le latin pour le français, où il pourrait puiser pour créer des doublets ou des formants. De nombreux termes arabes formés par recours au système de communication sont alors des équivalents de termes français (et anglais) qui le sont grâce à la composition savante, procédé très peu fréquent en arabe. Ils en sont le développement syntaxique.

Par exemple, l'un des équivalents arabes de *photoélectricité* est *kahrabâ ?daw ?iyya* («électricité lumineuse»)^{18,19,20}.

¹⁵ De plus, dans cet exemple, il y a trois termes *polarisation* : l'un concerne l'électromagnétisme, l'autre l'électrocinétique et enfin, le troisième, l'optique. Le terme arabe n'a fait ici que suivre les tribulations du terme français (ou anglais).

¹⁶ Pour ce qui concerne les termes eux-mêmes, nous savons, par exemple, comment les instances terminologiques veillent à éviter les mots tabous. Dans le domaine arabophone, ils peuvent de plus varier d'une région à une autre.

¹⁷ Ce type de connotation est du domaine de la psycholinguistique et devrait faire l'objet de tests.

¹⁸ La photoélectricité est le phénomène électrique d'origine lumineuse. L'on trouve aussi, comme équivalent, le terme de *daw? kahrabâ?iyy* («lumière électrique»), ce qui n'est pas la même chose.

¹⁹ Le terme *kahrabâ?* vient du persan *kahrobâ* («ambre jaune, succint»). C'est un emprunt ancien. L'on peut comparer cela au terme français *électricité*, formé sur le grec *elektron* («ambre jaune, succint» - *Dictionnaire Grec-Français* de Bailly). Quant aux Persans, pour *électricité*, ils utilisent le terme *barq*, emprunté... à l'arabe («éclair»; le télégramme se dit en arabe *barqiyya*). On a là l'exemple d'un bel échange terminologique ! Selon Vincent Monteil (1960 : 33:134), c'est Rifâ'a Râfi` at-Tahtawî (1801-1873) qui a le premier utilisé ce terme, en 1834, dans sa relation de voyage en France, *Tahîlîs al-Ibrîz ?ilâ talhîs Bârîz* (voir la traduction française *L'or de Paris.- Relation de voyage, 1826-1831*, par Anouar Louca, Paris, Sinbad, 1988 : 123).

²⁰ D'autres synonymes existent, qui font recours à des formants créés à cet effet, tels que *dawkahrabâ?iyya* et... *kahradaw?iyya*.

Si, en français ou en anglais, la facture de lexies à l'aide de formants grecs ou latin, indique d'emblée que ce sont — très probablement — des termes scientifiques, le développement de ces éléments dans le cadre de la syntaxe arabe, avec des éléments qui ne sont pas forcément des termes, est bien moins parlant. Ainsi *qarīb min al-mihwar* «proche de l'axe», équivalent arabe de *paraxial* est beaucoup moins marqué de ce point de vue²¹ que le terme français, ou que son synonyme *bârâmiḥwariyy*, créé, lui, avec le formant emprunté *bârâ-*.

Les unités terminologiques complexes, plus longues que les unités simples, sont susceptibles de contenir davantage d'information, d'être plus motivées que celles-ci. Mais cela n'est qu'apparence, n'enlevant rien au caractère réducteur de la dénomination. Ainsi l'exemple ci-dessus de *kahrabâ ? daw ?iyya* ne fait sens que grâce à la connaissance du référent.

c) le recours aux transferts sémantiques

Nous avons évoqué plus haut le fréquent réemploi de lexies existant déjà. Nombreux sont les termes d'optique formés par recours à la métonymie, à la métaphore et à l'hypallage. La plupart du temps l'on retrouve les mêmes glissements de sens qu'en français ou en anglais²².

Ainsi, en ce qui concerne la métaphore, l'équivalent arabe de *lentille* ('*lens*' en anglais pour la pièce optique, et '*lentil*' pour le légume) est *ʿadasa* : ce terme est inconnu en optique chez les savants arabes du Moyen Âge²³. Il s'agit bien du légume bien connu. Ce terme a en Europe une histoire intéressante (Vasco Ronchi, 1966 : 29-30) : la découverte des verres de lunettes «se produisit entre 1280 et 1285, très certainement dans la vallée de l'Arno», probablement par quelque artisan vitrier, par hasard, qui permettaient de corriger la presbytie :

«À cette époque, les scientifiques, bien que connaissant cette affection (le nom en est grec et très ancien), ne savaient pas du tout à quoi elle était due; non plus d'ailleurs que l'effet optique des disques de verre à faces convexes. Ces disques furent appelés *lentilles de verre* à cause de l'analogie de forme qu'ils offraient avec les lentilles comestibles. Un tel nom est une autre preuve de l'origine artisanale de cette invention. Jamais, à l'époque, un scientifique n'aurait donné le nom d'un légume à l'une de ses découvertes».

²¹ Dans ce syntagme, seul l'élément *mihwar* (*axe*) est lui-même un terme. Un autre équivalent arabe existe aussi pour *paraxial*, c'est *mutamahwir*, formé dans le cadre du système de nomination sur la racine quadriconsonantique (m - h - w - r), elle-même créée à partir du terme *mihwar*, lui-même de racine triconsonantique (h - w - r) : c'est un exemple de création de racine à partir d'un terme.

²² Voir sur cette question Lelubre (1992, chapitre 3.3).

²³ Ibn al-Hayṭam, parmi d'autres savants, avait fait des expériences, des montages où des lentilles étaient utilisées, mais il n'a pas de termes pour les désigner. Cependant, Roshdi Rashed (1993 : 233-234), à propos du terme *ballâr* ou *billâr*, cristal de roche, emprunt, avec métathèse, du grec *bêrullos*, évoque un passage du minéralogiste at-Tifâsî (m. 1253) et un autre, d'un astronome, Taqî ad-Dîn b. Ma`rûf (ouvrage achevé en 1574), où le terme *ballâra* est utilisé, qui semble indiquer une lentille plan-convexe, en cristal de roche.

Quant aux lentilles à faces concaves, comme «aucune à faces creuses n'existait parmi les légumes, les verres pour myopes ne furent pas nommés *lentilles* mais *verres creux* ».

Pour ce qui est de la métonymie, nous avons, par exemple, le cas de *ad-daw* ?, la lumière, employé pour *al-basariyyât*, l'optique, comme en français, et de même *su`â`daw* ?*yy*, rayon lumineux employé comme *su`â`basariyy*, rayon optique.

Le recours aux éponymes représente un cas particulier et fréquent de métonymie; les éponymes constituent un lieu de mémoire flagrant. Ainsi nos bien françaises *lois de Descartes*²⁴, ailleurs *lois de Snell* (ou de *Snellius*), ou encore, plus oecuméniques, *lois de Snell-Descartes* deviennent en arabe, selon que prédomine l'influence française ou l'influence anglo-saxonne — c'est-à-dire que sont adoptés les points de vue français ou anglo-saxons —, *qânûn-â Dikârt*, *qânûn-â Snîl*, et dans certains ouvrages, elles deviennent carrément arabes : c'est *qânûn-â Ibn al-Haytam*, les *lois d'Ibn al-Haytam*²⁵ !

Comme exemple d'hypallage, le réseau *sinusoïdal*, *muḥazzaza zaybiyya*, n'est pas un réseau de diffraction dont la forme serait sinusoïdale, mais en fait un réseau dont la périodicité est sinusoïdale.

d) l'emprunt

L'emprunt, bien sûr, indique la provenance étrangère. Il peut concerner une unité terminologique entière ou il peut être seulement celui d'un formant. Il peut aussi d'une part permettre de différencier des sources étrangères et d'autre part de donner des indications sur les procédés de la naturalisation du terme dans la langue d'accueil. Prenons deux exemples en arabe.

L'emprunt nous en dit aussi encore davantage, si l'on considère la façon dont il se fait sur le simple plan phonétique. Ainsi à côté de *mîkrûskûb* (sur la prononciation française), l'on trouve, *mâykrûskûb* (prononciation anglaise).

Évocatrice est aussi la transcription de certains noms de savants, qui n'étaient ni francophones ni anglophones. Ainsi *Kirchhoff* (1824-1887), allemand et germanophone, devient en arabe soit *Kîrsûf* soit *Kîrtsûf*, selon la prononciation française ou anglaise. Ces transcriptions ne sont-elles pas révélatrices d'une vision, induite par l'histoire contemporaine, francophone ou anglophone du reste du monde ?

²⁴ On disposait auparavant de tables de réfraction, grâce notamment à Witelo. Selon Bernard Maitte (1981 : 70), il semblerait que Descartes (sa *Dioptrique* date de 1637) ait eu directement connaissance des travaux de Snell, décédé prématurément, qui aurait trouvé les relations exactes entre angles d'incidence et de réfraction vers 1625.

²⁵ La loi de la réflexion (égalité de l'angle d'incidence et de l'angle de réflexion) ne posait pas problème. Il n'en était pas de même pour la réfraction, pour laquelle si la mise au point de tables précises de réfraction a été faite chez les Arabes — Ibn al-Haytam en a établies —, l'établissement d'une relation mathématique était autre chose.

La graphie arabe utilisée²⁶ peut être parlante : ainsi le phonème non arabe /g/, que toutes les institutions interarabes recommandent de transcrire par le graphème arabe *gayn*, est généralement rendu par le graphème *zîm* en Égypte (où il est prononcé [g]); le phonème /z/, quant à lui, est rendu par un graphème de même forme, mais comportant trois points au-dessous au lieu d'un seul; mais ce graphème, avec trois points dessous est utilisé en Irak, à l'instar des Iraniens, pour représenter le phonème /ts/, par le graphème *kâf*, surmonté d'un trait oblique (à l'iranienne, en Irak) ou de trois points (au Maroc), ou encore, comme souvent en Tunisie, par le graphème *qâf*, surmonté ou non de trois points²⁷, diacritisation qui à elle seule, évoque une différence entre le Moyen-Orient et le Maghreb sur le plan calligraphique²⁸.

L'emprunt peut s'effectuer selon différents degrés d'intégration. Ainsi pour le terme *fluorescence*, on trouve en arabe deux équivalents : *fulûriyya* et *tafalwur*. Le premier est formé par l'adjonction du suffixe *-iyya* sur l'emprunt *fulûr*, ou *filûr*, voire *fallûr* ou *falwar*; ce suffixe a de nombreux emplois, dont celui qui correspond aux suffixes *-escence*, *-ité*, *-isme* du français. Le second, quant à lui, est aussi formé à partir de l'emprunt *fluor*, mais de manière bien différente : de l'emprunt a été extraite une racine quadriconsonantique, la racine (*f - l - w - r*), sur laquelle ce terme a été formé, grâce au système de nomination de l'arabe, selon le schème [*taR₁aR₂R₃uR₄*], *R₁* étant la première consonne radicale, etc. De ce fait c'est un terme de facture arabe, formé sur une racine, elle, créée à partir d'un emprunt²⁹.

Des différents exemples évoqués ci-dessus, il apparaît que l'on peut classer les termes en deux catégories : ceux pour lesquels la forme nous donne des indications qui dépassent le cadre de la terminologie (présence d'éponymes, emprunts, formations savantes pour les termes français et anglais; présence de formants ou d'affixes spécifiques), et les autres, dont la forme — morphologie, syntaxe — ne les distingue en rien des lexies de la langue commune³⁰. Pour le premier groupe les termes présentent de manière intrinsèque des signes extra-terminologiques. Ce n'est qu'une approche extérieure qui peut nous renseigner sur des faits extra-terminologiques concernant les termes de l'autre groupe.

²⁶ En l'absence de convention véritablement respectée dans tous les pays arabes de transcription en caractères arabes de phonèmes non arabes.

²⁷ Ou tout simplement le graphème *qâf*, prononcé [g] dans de nombreux dialectes, dans le Monde arabe.

²⁸ Il s'agit de la différence de diacritisation des graphèmes *qâf* et *fâ*?

²⁹ Ce procédé est très ancien et explique la création de bon nombre de racines quadriconsonantiques en arabe. Les racines créées de nos jours, pour des besoins terminologiques, sont majoritairement de ce type.

³⁰ Par exemple il n'existe pas en arabe de doublet savant d'une lexie courante, qui serait dû à une évolution phonétique différente (comme cela arrive en français), ou de doublets de racines consonantiques (comme c'est le cas pour l'hébreu, qui peut faire recours à des racines araméennes). En revanche, les racines arabes formées à partir de lexies arabes ou d'emprunts, elles, sont porteuses de nouveauté par rapport au stock des racines de la langue.

3. LA CHARGE DE MÉMOIRE D'UN TERME

3.1 Les types de relations qu'un terme peut avoir

L'on peut se demander quels sont les types de relations (au sens le plus large) qu'un terme peut entretenir avec certains types d'entité, relations susceptibles d'être porteuses de mémoire. Deux types s'imposent :

(a) les relations référentiellement induites

- le terme avec l'unité référentielle qu'il dénomme (en particulier sa motivation);
- le terme avec les autres termes du «système terminologique» dont il dépend;
- le terme dans des rapports d'hyponymie, comme composant d'autres termes, dans des rapports de synonymie;
- le terme utilisé dans un autre domaine;
- le terme déterminologisé, entrant dans la langue commune.

Ces relations peuvent varier avec l'évolution du domaine.

(b) les relations de type lexical

- la facture du terme (ses composants, ses formants, la façon dont il a été formé, les phénomènes de calque,...);
- avec d'autres lexies de la langue;
- la connotation du terme.

3.2. Conclusion : aplatissement et distorsion du «temps»

Dans bien des domaines, la terminologie apparaît comme partiellement incohérente; cela est vrai si on la considère d'un point de vue strictement synchronique, si l'on fait abstraction de son histoire, de sa construction au cours du temps : il y a alors, dans cette façon de considérer les faits, ce que l'on pourrait appeler un *aplatissement* du «temps». Le temps, pour les terminologies, nous l'avons vu, c'est celui qui commande l'évolution, c'est-à-dire la construction, l'organisation, la réorganisation en fonction de nouveaux éléments de chacun des domaines concernés — comme nous l'avons évoqué pour l'optique —, évolution non linéaire.

Une telle vue strictement synchronique élimine, bien entendu, tous les termes qui ont disparu, soit en raison de la réorganisation du domaine — disparition d'unités référentielles — ainsi, *l'éther* du siècle dernier, depuis l'expérience de Michelson et Morley —, soit en raison de changements, de rectifications terminologiques, comme par exemple, dans la terminologie syrienne de la physique, pour *l'énergie*, le terme *qudra*, remplacé par le terme *tâqa*, utilisé partout ailleurs.

Mais, dans le cas de l'arabe en particulier, où coexistent, synchroniquement des variantes terminologiques, la prise en compte des variations régionales est par contre inévitable.

Ce que nous pouvons appeler *distorsion* du «temps» correspond à l'évolution différentielle au cours du temps d'un domaine par rapport à un autre, avec des phénomènes d'inertie terminologique, d'où les phénomènes de distorsion entre ou au sein de terminologies.

RÉFÉRENCES

TERMINOLOGIE, LINGUISTIQUE

REY, Alain (1979) : *La terminologie : noms et notions*, Paris, PUF, 1^{re} éd., 128 p.

RONDEAU, Guy (1993) : *Introduction à la terminologie*, Québec, Gaëtan Morin, 2^e éd., 238 p.

ARABE

LELUBRE, Xavier (1992) : *La terminologie arabe contemporaine de l'optique : faits - théories - évaluation*, Thèse de Nouveau Doctorat, Université Lyon 2, 546 p.

MONTEIL, Vincent (1960) : *L'arabe moderne*, Paris, Klincksieck, 386 p.

ROMAN, André (1990) : *La grammaire de l'arabe*, Paris, PUF, coll. «Que sais-je ?», 128 p.

OPTIQUE

MAITTE, Bernard (1981) : *La lumière*, Paris, Seuil, coll. «Points», 345 p.

NAZÎF, Mustafâ (1942-1943) : *al-Ḥasan b. al-Hayṭam : buḥûtu-hu wa kuṣûfu-hu [// Ibn al-Hayṭam : recherches et découvertes //]*, 2 volumes, Le Caire, Matba`a Nûrî, 879 p.

RASHED, Roshdi (1970) : «Optique géométrique et Doctrine optique chez Ibn al-Haytham», *Archive for History of Exact Sciences*, vol. 6, n°4, pp. 271-298.

RASHED, Roshdi (1993) : *Géométrie et dioptrique au X^e siècle — Ibn Sahl, al-Qûhî et Ibn al-Haytham*, Paris, Les Belles Lettres, 315 + 7 p.

RONCHI, Vasco (1966) : *L'optique, science de la vision*, Paris, Masson, 158 p.

SIMON, Gérard (1989) : «L'optique d'Alhazen et la tradition ptoléméenne : une nouvelle insertion dans le champ du savoir», *Colloque International d'histoire des sciences et de la philosophie arabes*, Paris, Institut du Monde Arabe, 22-25/11/1989, 7 p.

DE L'EMPLOI LIBRE À L'EMPLOI SUPPORT

Hassane FILALI SADKI

Université de Franche-Comté, Besançon, France

Le processus qui conduit un élément lexical — verbe, nom, préposition,... — à devenir un support de prédication reste encore un mécanisme mal connu et pas assez étudié. Nous essayerons de voir comment s'effectue le processus de délexicalisation d'une entité donnée qui perd soit totalement, soit partiellement sa valeur pleine pour ne devenir qu'un élément presque transparent à la relation prédicative. Toutefois, on constatera que dans cette perte significative de la valeur lexicale de base, l'élément qui devient support de prédication d'un N prédicatif garde une empreinte de sa fonction lexicale originale.

Notre hypothèse est que l'élément lexical, candidat à la fonction support, se vide dans son parcours de sa valeur élémentaire. Il a été considéré, jusqu'à présent, dans les différents travaux, que c'est le verbe qui sélectionnait les substantifs avec lesquels il se combinait. Or, les travaux récents, menés autour de M. Gross au LADL et G. Gross au LLL, démontrent plutôt l'inverse. Il s'est avéré que c'est les N prédicatifs qui sélectionnaient les verbes supports. Ces derniers n'ont comme fonction, en plus de celle d'être supports de la prédication et actualisateurs du temps, du nombre et de la personne, d'exprimer une certaine modalité d'action et l'aspect. Cette modalité d'action ne leur est pas inhérente dès la base lorsqu'ils sont employés dans leur valeur lexicale pleine d'une part et que les verbes qui ont cette particularité constituent une classe très réduite. Chaque item prend alors une valeur «modalisatrice» particulière en fonction du nom prédicatif qui le sélectionne.

Nous verrons par la suite que l'une des particularités des supports est qu'ils télescopent et figent une partie de leur construction, de sorte que toute modification de la structure syntaxique devient fatale à la grammaticalité de la phrase. Ainsi, si l'on prend pour exemple le verbe *prendre*, nous constatons qu'il y a une distorsion entre sa valeur lexicale pleine en tant que verbe d'action exprimant un processus et sa valeur délexicalisée où il est support.

L'étude des emplois spécifiques de *prendre*, en tant que verbe support ou opérateur, n'est possible qu'en faisant une analyse de toutes les structures syntaxiques dans lesquelles il s'insère, et l'analyse de tous les substantifs qui peuvent lui être associés. Notre étude portera essentiellement sur les emplois libres, les emplois supports et en fin les emplois figés. Nous nous interrogeons sur les mécanismes qui permettent à la même entité lexicale

d'avoir des emplois aussi variés et diversifiés et comment on peut trouver un lien entre eux tout en justifiant les changements à la fois syntaxiques, sémantiques et fonctionnels. Ainsi, pour éviter des confusions, il est nécessaire de spécifier les propriétés de chaque construction. Cette nécessité d'éclaircissement s'impose dans la mesure où le lexique n'est pas monosémique et que les verbes supports ont, le plus souvent, des valeurs différentes.

EMPLOIS INTRANSITIFS ET PRÉPOSITIONNELS

- (le vaccin + le feu + cette plante + le mortier + le plâtre) a bien pris

Ce type de phrase présente des propriétés syntaxiques particulières, et leur constructions diffèrent de celles que nous analysons pour *prendre*. Tout d'abord, ces constructions en *prendre* n'ont pas de compléments prédicatifs, ce qui est l'objet de cette étude. D'autre part, pour certains types de ces constructions, il est possible d'avoir une construction factitive.

- Luc fait prendre (le feu + la mayonnaise)
- (le feu + la mayonnaise) prend

Par contre, des constructions similaires à celles-ci font l'objet de notre étude.

- la teinture prend dans ce tissu
- ce tissu prend la teinture

Il en est de même de :

- Le feu prend dans cette maison
- cette maison prend feu

On peut dire que, dans la perspective du «lexique-grammaire», les entrées du dictionnaire ne sont pas des mots isolés, mais plutôt des phrases. Par conséquent, il y a autant de verbes, pour une unité définie morphologiquement comme telle, que de constructions transformationnellement possibles

EMPLOIS CONCRETS

Les emplois concrets de *prendre* acceptent la structure à trois arguments N₀ prendre N prép N₁. Les arguments de *prendre* peuvent être isolables à l'aide de questions :

- 1- Luc a pris un livre à Max
- qui a pris un livre à Max ?
- Luc
- qu'a pris Luc à Max ?
- un livre
- à qui Luc a-t-il pris un livre ?
- à Max

Ce test de questionnement permet de spécifier que $N_0 = N_{\text{hum}}$, $N_1 = N$ concret et $N_2 = N_{\text{hum}}$. Notons que le complément N_2 n'est pas nécessairement un N_{hum} , comme dans le cas de l'exemple suivant:

2- Luc a pris un livre à la bibliothèque

Dans ce cas, la question appropriée est où :

- où Luc a pris le Livre ?
- à la bibliothèque

Lorsque le verbe *prendre* a un complément concret, il a un emploi libre, et de ce fait il est possible de lui substituer d'autres verbes.

- Luc a (emporté + emprunté + volé + feuilleté) ce livre

Dans cette structure, le verbe *prendre* exprime un changement de localisation dans l'espace, puisqu'il y a un avant et un après. Ainsi, pour (1) et (2) on peut dire que avant : Max avait le livre ou qu'il (était + se trouvait) à la bibliothèque et qu'après : Max n'a plus le livre ou que le livre n'est plus à la bibliothèque

PRÉSENTATIONS DES LISTES

Les substantifs représentés dans les listes sont des prédicats nominaux qui forment avec le verbe support *prendre*, dans une phrase simple, une entrée lexicale autonome. Ces N prédicatifs sont répartis dans différentes tables en fonction de la nature du substantif et de celle du déterminant. Ainsi nous avons soit des N libres (à déterminants non contraints) soient des N non libres (à déterminants figés). À l'intérieur de cette bipartition, les N prédicatifs sont classés d'après leurs constructions syntaxiques.

Dans les listes à déterminant figé, on a remplacé le symbole N par le symbole C pour indiquer la présence d'un élément figé. Étant donné que le déterminant est figé, on peut dire qu'il fait partie du prédicat nominal. C'est la raison pour laquelle, l'élément C représente à la fois le N prédicatif et son déterminant. Ces constructions figées sont réparties en tables en fonction de leur structure syntaxique. À l'intérieur de chaque structure, nous avons établi des sous-structures en fonction de la nature de l'élément nominal, du modifieur : adjectival ou prépositionnel.

PROBLÈMES GÉNÉRAUX DE SYNTAXE ET DE LEXIQUE

Les prédicats nominaux que nous étudions sont classés en fonction de leur construction, et plus particulièrement en fonction de la nature de leur argument. La première démarche consiste à distinguer d'abord les N prédicatifs qui prennent un complément de ceux qui sont en construction intransitive. Les compléments des V-n sont généralement de forme prép N_1 . Ainsi, en fonction de ce critère on peut avoir les différentes structures où les N prédicatifs ont pour support le verbe *prendre*.

- (1) N_0 prendre (E + Dét) N
- Luc prend l'air

- la maison prend feu

(2) N₀ prendre (dét N) (prép N₁)

- Luc prend Max sous sa direction

(3) N₀ prendre (Dét N) (à N₁)

- Luc prend cette proposition au sérieux

(4) N₀ prendre (dét N) (de N₁)

- Luc prend des informations de Max

(5) N₀ prendre N₁ (prép N)

- Luc prend Max en charge

- Luc prend Max en haine

Donc, le critère de la complémentation nous permet d'établir une classification provisoire, et d'avoir trois types de constructions, celles présentées en (1) sans complément, la seconde comprend un complément prépositionnel, les cas de (2) à (4). Ces trois structures ont en commun le fait que le prédicat nominal est en construction directe. Tandis que dans la construction présentée par la structure (5), le prédicat nominal est en construction prépositionnelle, tandis que le complément est en construction directe. Les formes à complément prépositionnel ont été, à leur tour, subdivisées en plusieurs classes.

PHRASE SIMPLE OU À COMPLÉMENT COMPLEXE ?

La question de savoir s'il s'agit de phrase simple ou de constructions complexes s'impose, presque de manière automatique, à chaque fois qu'il est question d'analyser une structure à verbe support de type N₀ V_{sup} N prép N₁. Ainsi :

1- Luc prend Max sous sa direction

2- Luc prend la décision de partir

3- Luc prend Max en croupe

On constate que lorsque le V-n se trouve en position de prédicat nominal, le cas de décision, est rattaché à un verbe, ici décider, le statut de chaque argument N₀ et N₁, est prédéfini par leur relation dans la construction verbale.

- Luc décide de partir

La question qui se pose est de savoir si les constructions avec le verbe support *prendre* lorsqu'elles peuvent être reliées soit à des phrases ayant d'autres supports : être, avoir, etc., présentent ou non entre elles des relations de dépendances par différents procédés syntaxiques. Nous avons vu, dans le cas du complément de N, qu'il était possible de relier les phrases entre elles par relativisation.

On voit que la réponse à cette question n'est pas aussi simple, et demande une attention particulière. D'autant plus que la diversité des compléments prépositionnels ne permet pas, au premier abord, de savoir s'il est question ou non de structure simple ou de

phrase complexe. Nous reviendrons sur cette question dans différentes parties de ce travail, et plus particulièrement lors de l'analyse des constructions prépositionnelles en *prendre*.

RELATION DE N À N₀ : FORMATION D'UN GN

Si l'on compare les phrases, on constate que le substantif *chapeau* peut avoir un complément de N_{hum}, mais non pas *décision* :

- Max prend (un chapeau + une décision)
- Max a pris (le chapeau de Chloé + *la décision de Luc)

Le substantif *chapeau* prend un complément de N_{hum} qui est différent du N₀, tandis que *décision* nécessite la coréférence de ce complément avec le sujet du verbe *prendre*.

- Max a pris sa décision

La possibilité ou non d'avoir un complément nominal est un critère pour savoir si le N₀ est à la fois le sujet du verbe support et du prédicat nominal ou non. Si l'on remplace *prendre* par un autre verbe comme *apprécier*, par exemple, il devient possible d'avoir un complément prépositionnel de N_{hum}.

- Luc apprécie (cette décision + la décision de Max)

Dans cette dernière phrase, *Luc* est le sujet du verbe *apprécier*, mais il ne peut pas être l'agent du prédicat nominal. D'ailleurs, il est possible de restructurer, dans ce cas, le verbe support *prendre*, par relativisation.

- Luc apprécie la décision qu'a prise Max

Parallèlement à cette construction, il est possible d'avoir *Max* en position de sujet du prédicat nominal à l'intérieur de la complétive enchâssée après le verbe *apprécier*.

- Luc apprécie que Luc prenne cette décision
- Luc apprécie le fait que Max ait pris rapidement sa décision

De la même façon, la relation du sujet entre N₀ et *prendre* peut être confirmée par la possibilité d'avoir la structure passive :

- Luc apprécie la décision prise par Max

où *Max* est en position de complément d'agent. On peut dire, donc, que le N₀ entretient une relation de sujet à la fois avec le support *prendre* et le N prédicatif. Cette relation est l'un des critères qui permettent de distinguer les N prédicatifs des noms non prédicatifs.

DISTINCTION ENTRE N PRÉDICATIFS ET N NON PRÉDICATIFS

Cette question s'impose, vu la diversité des constructions dans lesquelles entre le verbe *prendre*. L'étendue du lexique dicte un choix à faire. S'il n'y a pas la moindre

hésitation à considérer des substantifs comme *décision* et *courage* comme des N prédicatifs, cependant, il est difficile de savoir ce qui distinguerait *train de chapeau*.

- Luc a pris (une décision + du courage + le train + le chapeau)

Nous avons vu que l'opposition N concret / N abstrait n'est pas un paramètre fiable pour une classification du lexique. Afin de bien distinguer les N prédicatifs des N non prédicatifs, il est nécessaire d'appliquer à l'élément nominal un ensemble de critères syntaxiques.

- prends (le livre + un morceau de tarte)
- *prends (le large + une gifle + plaisir)

D'autant plus, dans la valeur lexicale pleine, le verbe peut être remplacé — lorsqu'il exprime un processus ou une action — par la proforme faire, alors que celle-ci est incompatible quand le verbe en question — devient support du N prédicatif — avec sa valeur délexicalisée.

- Luc (mange + prend) une pomme et Max en fait autant
- *Luc prend (une gifle) et Max fait autant

En outre, lorsqu'un verbe est employé comme support, il permet même d'actualiser des N concrets par métonymie et qui deviennent de ce fait des vrais N prédicatifs :

- Luc a pris un verre = Luc a bu un verre
- Luc a pris son repas = Luc a mangé son repas

Même s'il est parfois difficile de mettre en relation les propriétés syntaxiques qui permettent l'actualisation de relations d'interdépendances entre certaines catégories grammaticales et les entrées lexicales, le choix d'une métonymie plutôt qu'une autre dépend en fait exclusivement de considérations extralinguistiques.

LES CONTRAINTES SUR LES DÉTERMINANTS

La détermination semble être un des critères qui permettent de distinguer les N libres des N non libres. Le mot *chapeau* dans la phrase suivante n'impose pas de contrainte particulière à sa détermination.

- Luc a pris (un + ce + le + mon + son) chapeau

Même le déterminant possessif qui vient de la structure N de N n'impose pas de contrainte de coréférence entre le sujet de *prendre* et le complément du nom *chapeau*. Son, par exemple, peut avoir deux référents différents dans cet exemple, puisqu'il peut être coréférent à Luc, comme il peut renvoyer à une tierce personne, Marie, par exemple. Contrairement à cela, dans une phrase à support, le N prédicatif impose la coréférence entre le possessif et le sujet du verbe *prendre*.

- Luc a pris (sa + *ta + *ma) décision

Cette contrainte sur la coréférence s'impose même lorsque N prédicatif n'émane pas nécessairement de N₀.

- Luc prend (ses + *tes + *mes) ordres de Max

Même si les ordres sont de Max, pourtant l'adjectif possessif doit avoir le même indice que le sujet N₀. Dans ce cas, il n'est pas possible, par exemple, d'avoir le déterminant défini *les* :

- *Luc prend les ordres de Max

LA QUESTION PAR *QUE* PORTANT SUR LE N₁

Le test du questionnement par *que* semble être déterminant pour distinguer les prédicatifs des N non prédicatifs. Pour les verbes pleins, le test question-réponse par *que* permet de savoir si N₁ assume la fonction de complément d'objet. Ainsi, pour le verbe *prendre* on a :

- Luc a pris son chapeau
- Qu'a pris Luc ?
- son chapeau

Par contre, dans son emploi de verbe support, le test question-réponse produit un énoncé erroné.

- Luc prend du plaisir
- Que prend Luc ?
- du plaisir

Luc prend ses ordres de Max
- Que prend Luc de Max ?
- ses ordres

Luc prend ce cours sous sa responsabilité
- Que prend Luc sous sa responsabilité ?
- ce cours

Luc prend l'air sur la terrasse
- Que prend Luc ?
- l'air sur la terrasse

LES DÉTERMINANTS DE N : ARTICLE INDÉFINI OU PAS

La question de la détermination est primordiale pour savoir si, dans la construction à verbe support, le prédicat nominal est un N libre ou s'il est question d'une structure figée. D'ailleurs, la possibilité ou non de la variation de la détermination est en relation avec l'application ou non d'un certain nombre de transformations syntaxiques. Ainsi, si nous considérons les phrases suivantes :

- Luc prend (un + ce + ton + son) chapeau
- Luc prend (un + un beau + un certain) livre de la bibliothèque

Le déterminant, du complément d'objet *chapeau* et *livre*, n'est pas contraint syntaxiquement. Cependant, ceci n'est pas généralement le cas avec les N prédicatifs dans les constructions à support. Nous avons vu que les constructions en *prendre* imposent par contre des contraintes à la détermination du prédicat nominal.

- Luc a pris une (E + fausse) information de Max

On voit bien que le substantif, *information*, après le support *prendre* peut aussi avoir des déterminants variés. Cependant, la contrainte qui pèse sur ce type de prédicats est la nécessité de coréférence entre le sujet N_0 de *prendre* et le complément prépositionnel N_1 . Ce dernier étant à son tour le sujet du N prédictif. Ceci est confirmé par le fait que le possessif avec le V-n doit renvoyer nécessairement à N_0 .

- *Luc prend Max sous (ma + ta) responsabilité
- *Luc prend Max sous (mes + tes) ordres

Cette propriété de coréférence est particulière à l'emploi des verbes supports par opposition à l'emploi des verbes ordinaires.

- Luc a acheté (mon + ton + son) journal
- Luc a mangé (ma + ta + sa) soupe

Nous avons vu que la contrainte de coréférence était l'élément essentiel pour la formation de complément N_{hum} avec un N prédicatif par opposition à un N non prédicatif.

- *Luc a pris la décision de Max
- Luc a acheté le journal de Max
- Luc apprécie la décision de Max

Le choix des déterminants possibles pour un complément est un critère essentiel qui est en relation avec la possibilité ou non de l'application de certaines transformations. Ainsi, on peut mettre en évidence la relation qui existe entre le fait qu'un N prédicatif prenne le déterminant indéfini et la transformation relative. Cette relation est validée par les constructions à support *prendre* que nous analysons. Ce lien entre la détermination et la possibilité de la transformation relative est extensible à tout le lexique.

- Luc prend conseil auprès de Max
- le conseil que Luc prend auprès de Max

Nous avons vu, auparavant, que la relativisation permettait la formation de GN de forme : N de N_0

- Luc est surpris du dessus que Max prend sur Chloé
- Luc est surpris du dessus de Max sur Chloé

Cependant, la formation de la relative n'est pas systématique, puisque certains substantifs se prêtent mal à une telle transformation, malgré le fait que le déterminant n'est pas totalement figé.

- Luc prend (un + un grand) plaisir à lire ce livre
- Le plaisir que Max prend à lire ce livre [est particulier]
- *Son grand plaisir à lire ce livre [est particulier]

Le fait que le N prédicatif accepte ou non le déterminant indéfini peut être un critère décisif dans la classification du lexique. En effet, il permet de distinguer les N libres, qui n'ont pas généralement de contraintes particulières sur la variation de leurs déterminants, des N non libres, et qui devient de ce fait une construction figée. La question de la variation de détermination est un facteur qui autorise ou non la formation de groupes nominaux de forme le N de N₀.

- Luc prend (E + un particulier) plaisir
- le plaisir que prend Luc
- son plaisir

Le prédicat nominal, *plaisir*, n'accepte ni le déterminant indéfini, ni le déterminant défini.

- *Luc prend (un + le) plaisir

Cependant, l'utilisation de l'article indéfini nécessite la présence d'un modifieur. Le fait que plaisir prenne le déterminant Un - Modif est le signe qu'il est question d'un substantif libre. La preuve en est la possibilité d'utiliser d'autres déterminants.

- Luc prend (beaucoup de + du) plaisir

Dans la présentation des tables, les déterminants indéfinis — un + Un - Modif, du + des — sont notés dans les colonnes. Ce qui permet de prendre en compte des cas où le déterminant indéfini est obligatoirement accompagné d'un modifieur. Ceci permet, par exemple, de distinguer les différents emplois, du prédicat.

- 1a - Luc prend l'air
- 1b - Luc prend de l'air
- 1c - *Luc prend un air
- 2a - Luc prend un air sévère
- 2b - *Luc prend l'air sévère

Dans (1a) et (1b), le substantif *air* renvoie à l'oxygène. D'ailleurs, dans ce cas, il est possible d'utiliser un quantifieur.

- 1d - Luc prend une bouffée d'air

Parallèlement, il est possible d'avoir le déterminant Du - Modif.

- 1e - Luc prend de l'air pur

1f - Luc prend une bouffée d'air pur

Or, lorsqu'il est question de l'apparence, il n'est pas possible d'avoir Du - Modif.

2c - *Luc prend de l'air sévère

De même, la quantification n'est plus possible.

2e - *Luc prend une bouffée d'air sévère

Pour ce qui est des constructions figées, nous les avons classées dans des tables particulières. Elles se caractérisent par le fait que le déterminant est invariable, et que chaque entrée lexicale ne prend qu'un déterminant spécifique. Ce caractère fait du Dét une partie intégrante du prédicat nominal. Pour ce type de tables, nous n'avons consacré qu'une seule colonne à Dét, et que nous avons noté en toutes lettres. Nous avons ainsi :

- Luc prend (le large + son pied + terre)

Cependant, pour les constructions figées, nous n'avons pas spécifié pour chaque déterminant s'il était ou non accompagné d'un modifieur. Ainsi, dans les tables, nous avons regroupé les N simples, (comme large + pied ...) et des N composés à modifieur adjectival AN ou prépositionnel N prép N. Ce paramètre d'extension du prédicat nominal a été pris en considération dans la subdivision des constructions figées, comme d'ailleurs pour les autres tables, en fonction de la nature du substantif. Ainsi, nous avons pour la table N *prendre* C :

- le cas où N est seul
- Le cas où N est accompagné nécessairement d'un adjectif
- Le cas où N est suivi obligatoirement d'un modifieur prépositionnel.

LES N LIBRES : LES CONSTRUCTIONS PRENDRE N SOURCES DE GN

Nous avons discuté auparavant des critères qui permettent de distinguer les N prédicatifs libres des N prédicatifs non libres. L'un des critères que nous avons retenu était la possibilité d'avoir ou non un déterminant indéfini, exception faite des constructions où il est figé.

- Luc a pris une gifle

Cette propriété fait que les prédicats nominaux à supports fonctionnent comme n'importe quel substantif dans n'importe quelle construction, puisqu'il peut être aussi bien en position sujet qu'en position complément.

On constate donc que le verbe *prendre* reçoit dans la structure de base ou libre une valeur de verbe plein, qui reste constante puisqu'elle dépend de son statut lexical. Cette valeur est déterminée par sa structure distributionnelle facilement définissable et qui correspond à sa structure la plus étendue ou chaque argument est utilisé pour ainsi dire dans son sens concret ou premier. Tandis que dans les emplois supports, le verbe, malgré

la parenté de la structure dans laquelle il s'insère, n'a presque plus rien gardé du sens premier, et qu'il s'est vidé presque de son sens, et acquiert des valeurs sémantiques différentes dépendantes des N prédicatifs auxquels il s'associe et dont il est le support. Dans ce processus d'extension d'utilisation, le verbe plein devient presque transparent à la relation prédicative, à l'image des opérateurs aspectuels. Cependant, il garde, en partie, souvenir de sa combinatoire libre.

Ainsi, dans ce processus de désémantisation et de délexicalisation, le verbe support même s'il voit modifier les contraintes distributionnelles qu'il imposait ou qui le reliaient à ses arguments dans les constructions libres, il ne rompt pas définitivement pour autant sur le plan syntagmatique et sémantique avec son passé historique et garde une certaine image de ses emplois.

Le plus souvent, et il n'y a pas très longtemps, on associait, de manière presque systématique, les verbes supports aux constructions contenant un nom d'action dérivé d'un verbe comme dans l'exemple suivant :

- Luc a décidé de partir
- Luc a pris la décision de partir

Ainsi, *prendre* est considéré comme un support de nominalisation, comme on parle dans d'autres cas de supports d'adjectivation. Cependant, on s'est rendu compte de l'existence de structures syntaxiques dans lesquelles le verbe assume le support de prédication d'un nom, sans que ce dernier puisse être relié de quelques manières, sauf par des relations ad hoc, à une base verbale ou adjectivale. D'autant plus que ces substantifs ne sont pas nécessairement des N abstraits. Ainsi dans :

- Luc a pris (le train + l'avion + le métro)

De même, l'existence de structures dont au moins un élément est figé :

- Luc a pris (le large + la mer)

Nous constatons alors l'existence d'un certain continuum entre les emplois libres, les emplois supports et les emplois figés. Ce continuum qui au départ n'était pas prévisible, mais grâce à l'étude systématique du lexique et à l'établissement de classe de noms, il est possible de justifier et d'expliquer les relations qui existent entre les différents emplois.

Ainsi, donc, grâce à ce continuum, on peut rendre compte à la fois de la diversité des emplois et des constructions syntaxiques, des conditions dans lesquelles ils acquièrent des sens diversifiés, puisque l'on peut dire qu'il existe autant d'entrées lexicales pour le même lexème. Cette diversité résulte à la fois du fait qu'il garde partiellement des traces sémantiques de son emploi libre, et en perd lorsqu'il contribue à la construction du sens des éléments prédicatifs auxquels il s'associe.

Ainsi, dans les emplois supports, l'élément en question en se délexicalisant, se neutralise pour n'être qu'un élément neutre de la prédication. Cette neutralité s'accroît dans le cas des constructions figées puisque l'élément en question perd totalement toute

aliénation avec son emploi libre comme dans le cas de l'expression *prendre le taureau par les cornes*.

Ce processus de neutralisation ne peut être déterminé de manière claire que dans le cadre du lexique-grammaire en analysant toutes les classes de noms, et les supports qui peuvent leur être associés. C'est au terme de ce travail qu'on pourrait rendre compte de la neutralisation des valeurs sémantiques de chaque mot, des phénomènes de la synonymie, des mécanismes de métaphorisation et de métonymie. Il serait intéressant de se pencher sur les phénomènes syntaxiques et sémantiques

La particularité de cette classe de mots — limitée en nombre et dont l'utilisation et la fréquence d'emploi sur le plan statistique est très importante — qui peuvent devenir des éléments neutralisés dans des emplois supports est de subir des changements sémantiques. Chaque mot entre son emploi premier, qui correspond à la définition sémantique du dictionnaire, l'emploi support et l'emploi métaphorique déploie toutes ses différentes significations comme s'il manifestait dans chaque construction l'une de ses différentes facettes. L'actualisation dans le discours d'un mot relève à la fois du même et du différent. Du même, puisque l'entité garde une trace pour ne pas dire un sème de son sens primaire comme s'il gardait en mémoire la raison de son existence ou la signification pour laquelle il a été créé et qu'il doit assumer. Du différent, puisque le processus de son actualisation dans un énoncé l'amène à des contraintes combinatoires, qui lui font perdre petit à petit et partiellement son sens jusqu'à le neutraliser complètement. À chaque fois qu'il se combine et s'associe à un N prédicatif, on a le sentiment qu'il perd ses propriétés syntaxiques et sémantiques puisqu'il constitue avec lui à chaque fois une entité nouvelle, ce qui a fait que plusieurs auteurs traitent les constructions supports comme des locutions ou expressions figées ou figurées. Certains parlent d'une coalescence verbo-nominale, puisque aussi bien le verbe que le nom prédicatif perdent chacun une partie de ses propriétés pour faire place tous les deux à une signification nouvelle qui n'est pas nécessairement la somme des significations de l'un et de l'autre. Chaque association, chaque actualisation renverse la ou les significations des items, ce qui rend difficile toute classification du lexique, vu le nombre des possibilités combinatoires et plus particulièrement des noms prédicatifs.

Les supports ont donc cette faculté de mutation et de transformation sémantique et syntaxique puisque leur entité mue et change systématiquement à chaque combinaison. Ils acquièrent des significations différentes à la fois par association à d'autres éléments lexicaux et par mimétisme et analogie puisqu'ils intègrent partiellement des significations de l'élément auquel ils sont combinés. Par conséquent, ils se mettent à ressembler à d'autres entités — généralement de la même catégorie et ayant la même fonction et parfois même différente, ce qui peut être relié au phénomène de synonymie — de sorte qu'ils peuvent assumer et prendre leur place dans un processus de substitution lexicale et actualiser au moins une des significations de ce dernier.

Les supports peuvent avoir entre autres comme fonction d'établir des relations facilement définissables et repérables avec les N prédicatifs qu'ils actualisent dans le discours. Ces relations peuvent se déduire de l'établissement de classes d'objets (Cf. G. Gross), ce qui permet de justifier et d'expliquer comment ils parviennent à actualiser l'une des significations de l'élément supporté, mais même si les résultats sont probants, il en reste que la distribution et le fonctionnement à l'intérieur de la même classe n'obéit pas de

manière systématique et mécanique aux mêmes règles combinatoires vu certaines comptabilités.

Comme on peut le constater, le figement de ces constructions ou de la séquence (verbe support - N prédicatif) incombe au fait des contraintes de détermination. Le type de déterminant, la possibilité ou non de le varier font qu'il devient un élément central de la construction puisque le figement dépend essentiellement de lui. C'est dans ce sens qu'on préconise la séparation des constructions figées des constructions à supports et des constructions libres. N'empêche qu'il y a un certain continuum entre les différentes constructions puisque «les règles que subissent les expressions figées sont exactement les règles de la syntaxe des phrases libres et ce aussi bien pour leurs parties libres que pour leurs parties figées» (M. Gross, 1988).

RÉFÉRENCES

- BOONS, J.-P., GUILLET, A. et Ch. LECLERE (1976) : *La structure des phrases simples en français : construction intransitive*, Genève, Droz, 377 p.
- GIRY-SCHNEIDER, J. (1978) : *Les nominalisations en français : l'opérateur faire dans le lexique*, Genève, Droz, 353 p.
- GIRY-SCHNEIDER, J. (1978) : *Les prédicats nominaux en français : les phrases simples à verbes supports*, Genève, Droz, 396 p.
- GROSS, G. (1989) : *Les constructions converses du français*, Genève, Droz.
- GROSS, G. (1994) : «Classe d'objets et traitement de la synonymie», Ibrahim, A. H. (dir.), *Supports, opérateurs, durée*, Annales de l'Université de Besançon 516. Série Linguistique et Sémiotique, vol. 23, Paris, Les Belles Lettres, 268 p.
- GROSS, G. (1994b) : «Classe d'objets et descriptions des verbes», *Langages*, 115, Paris, Larousse.
- GROSS, M. (1975) : *Méthodes en syntaxe*, Paris, Hermann.
- GROSS, M. (1976) : «Sur quelques groupes nominaux complexes», *Méthodes en grammaire française*, Paris, Klincksiek.
- GROSS, M. (1981) : «Les bases empiriques de la notion de prédicat sémantique», *Langages*, 63, Paris, Larousse.
- GROSS, M. (1982) : «Une classification des phrases figées du français», *Revue Québécoise de Linguistique*, XI-2, Montréal, Presses de l'Université de Québec.
- GROSS, M. (1986) : «Les nominalisations d'expressions figées», *Langue française*, 69, Paris, Larousse.
- GROSS, M. (1988) : «Les limites de la phrases figée», *Langages*, 90, Paris, Larousse.
- HARRIS, Z. S. (1970) : *Papers in Structural and Transformational Linguistics*, Dordrecht, D. Reidel.

- IBRAHIM, A. H. (1984) : «Sur le statut de quelques accidents syntactico-sémantiques», De la syntaxe à la pragmatique, vol. 8 de *Linguisticae Investigationes Supplementa*, Amsterdam/Philadelphia, John Benjamins.
- IBRAHIM, A. H. (1993) : «La déviance de la suffixation en français est-elle structurelle ?» *TRANEL*, Traitement des données linguistiques non standard, 20, Neuchâtel, Université de Neuchâtel.
- IBRAHIM, A. H. (dir) (1994) : *Supports, Opérateurs, Durées*, Annales de l'Université de Besançon 516, Série Linguistique et Sémiotique vol. 23, Paris, Les Belles Lettres, 268p.

LE PROJET NADIA-DEC : VERS UN DICTIONNAIRE EXPLICATIF ET COMBINATOIRE INFORMATISÉ ?

Gilles SÉRASSET

GETA-CLIPS-IMAG (UJF & CNRS), Grenoble, France

INTRODUCTION

Dans le domaine de l'ingénierie linguistique et de la connaissance, le problème des ressources lexicales et linguistiques s'est toujours posé. Néanmoins, l'avancée des techniques du Traitement Automatique des Langues Naturelles (TALN) l'a rendu plus sensible. Il nous faut maintenant pouvoir répondre à des besoins importants en termes de quantité, de qualité et de complexité. La complexité et la diversité des informations requises augmentent avec les exigences des outils de TALN ainsi qu'avec le développement de nouvelles applications (humaines ou machinales). Si la récupération (semi)automatique d'information lexicale est une piste, elle ne pourra remplacer la création manuelle de dictionnaires.

Nous nous sommes donc intéressé à la construction d'outils pour lexicographes et lexicologues. Afin d'avoir une bonne compréhension des problèmes qui se posent, nous avons décidé d'informatiser un dictionnaire complexe, contenant de nombreuses informations structurées, le *Dictionnaire explicatif et combinatoire du français contemporain* (DEC). Le DEC étant un travail de lexicologie, il ne s'agit donc pas à proprement parler d'un dictionnaire, mais plutôt d'un ensemble d'entrées destinées à illustrer une théorie linguistique. Ce ne sont donc pas les données que l'on a informatisées, mais le processus de rédaction de ces données.

Les travaux menés au cours du projet NADIA-DEC s'appuient d'une part sur le système SUBLIM (Sérasset 1994) défini au laboratoire GETA-CLIPS de l'Université Joseph Fourier (Grenoble I) et, d'autre part, sur les travaux de lexicologie menés par l'équipe d'Igor Mel'cuk (Mel'cuk et al. 1995) au laboratoire GRESLET de l'Université de Montréal. Il a reçu le soutien du réseau LTT de l'AUELF-UREF et des ministères français et canadiens des affaires étrangères.

Ce projet pose des problèmes informatiques et linguistiques sérieux. L'aspect évolutif de la structure du DEC impose la nécessité de fournir des outils informatiques adaptables. La nécessité de formalisation des informations peut conduire à différentes stratégies de représentation des informations.

Nous montrons les différentes étapes de l'informatisation du DEC en donnant tout d'abord la structure interne des informations lexicales. Nous donnons ensuite un aperçu des outils et méthodes utilisés pour la création et la validation d'un DEC informatisé.

LE PROJET NADIA-DEC

Objectifs

Depuis 1994, le GRESLET (Université de Montréal) et le GETA-CLIPS (Université Joseph Fourier - Grenoble I) travaillent ensemble sur le projet NADIA-DEC, soutenu par le réseau LTT de l'AUPELF-UREF.

L'objectif de ce projet est l'informatisation du *Dictionnaire explicatif et combinatoire du français contemporain* (DEC) créé par Igor Mel'cuk. Cette informatisation se base sur les travaux préalablement effectués au GETA (Sérrasset 1994), et répond aux contraintes suivantes :

- **Fidélité linguistique** : les structures informatiques utilisées doivent rester proches des structures linguistiques que l'on souhaite représenter;
- **Généricité** : les outils construits doivent pouvoir être utilisés pour d'autres structures;
- **Adaptabilité** : les outils informatiques doivent pouvoir évoluer en même temps que la structure informatique.

Le DEC étant en constante évolution, il nous est très vite apparu important d'informatiser non seulement les données existantes, mais surtout sa production. Nous avons donc développé des outils d'éditeurs spécialisés pour le DEC ainsi que des outils de récupération des informations déjà décrites disponibles sous forme de fichiers Word™.

L'approche utilisée ne remet pas en cause la structure linguistique que l'on peut trouver dans le DEC. La structure informatique du DEC doit permettre, au minimum, de re-générer à l'identique les fichiers Word™ utilisés pour la version papier. Aussi, toutes les informations sont présentes, et ce même si elles ne sont pas structurées. Il est ainsi toujours possible, au fur et à mesure que l'on avance dans ce projet, d'augmenter la structuration des données sans avoir à reprendre l'ensemble du processus de récupération à partir des fichiers Word™.

À cet égard, le projet NADIA-DEC se distingue des autres projets d'informatisation du DEC, qui se basent a priori sur une structure informatique simplifiée et qui n'informatise que le sous-ensemble de données commun entre le DEC et cette structure.

Enfin, les données du DEC ne sont pas récupérées dans le but d'une utilisation informatique particulière. Nous estimons que cette indépendance par rapport à l'usage qui sera fait des données nous permet de garantir la complétude des informations récupérées.

Méthodologie

Nous avons distingué plusieurs tâches pour accomplir le projet NADIA-DEC :

- définition d'une structure informatique pour le DEC,
- récupération des informations existantes sous cette forme structurée,
- construction d'un éditeur spécialisé pour cette structure (l'éditeur DECID),
- exportation des données structurées vers différentes formes.

Ainsi, notre méthodologie peut être résumée par le schéma suivant :

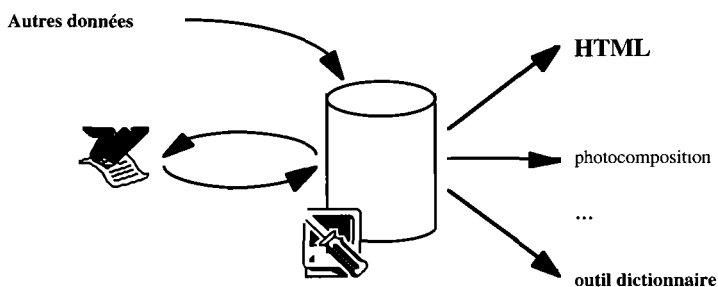


Figure 1 : Méthodologie de création d'un DEC informatisé

L'ÉDITEUR DECID

DECID est un éditeur spécialisé pour l'édition du DEC. Il offre de nombreuses fonctionnalités pour aider le lexicographe. Sa conception et son implantation ont été effectuées dans un souci de simplicité et de convivialité.

En utilisant l'éditeur DECID, le lexicographe crée ou modifie, en direct, une structure informatique. Pourtant, l'interface a été conçue pour lui donner l'impression de travailler, comme auparavant, sur le DEC tel qu'il est publié. Aussi, la visualisation des données est-elle très proche de celle qui est utilisée dans la version papier.

Le lexicographe dispose d'une fenêtre principale lui donnant la liste des vocables et des lexies du fichier en cours d'édition (figure 2). Le second type de fenêtre présente et permet d'éditer une lexie.

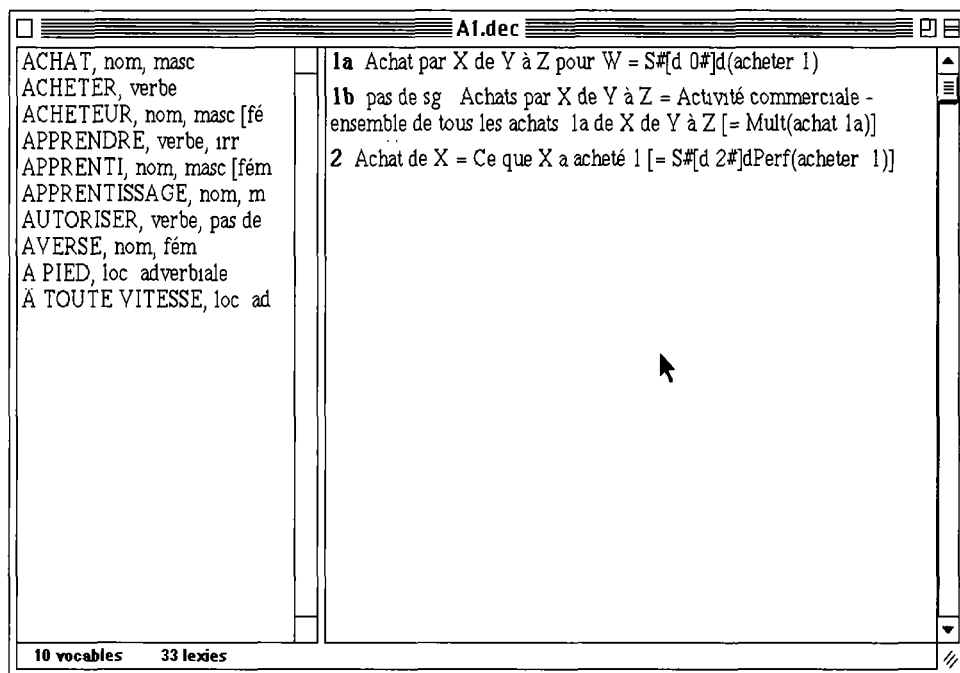


Figure 2 : La fenêtre principale. La zone de gauche présente la liste des vocables du fichier en cours d'édition. La zone de droite présente la liste des lexies du ou des vocables sélectionnés

La fenêtre de lexie permet d'éditer le vocable, le numéro de la lexie, les informations morphologiques, la définition et les exemples de manière très simple. La zone de régie n'est pas encore traitée par l'éditeur DECID.

Les fonctions lexicales apparaissent sous une forme très proche de la forme papier. Mais leur édition a été rendue très aisée par l'éditeur DECID. En effet, auparavant, le lexicographe devait, pour éditer une fonction lexicale, utiliser des indices, des exposants, des changements de fontes. Il devait faire attention à la correction du nom de la fonction, bien mettre le premier caractère en majuscule et le reste en minuscule... Tous ces soucis ont maintenant disparu lorsqu'on utilise l'éditeur DECID. En effet, le lexicographe se contentera de taper : perm1incepreal3+usual pour voir se dessiner la fonction : **Perm₁IncepReal₃^{usual}**.

Le lexicographe pourra ensuite sauver son travail sous forme structurée ou alors sous forme d'un fichier RTF (Rich Text Format) qu'il pourra ensuite utiliser directement en Word.

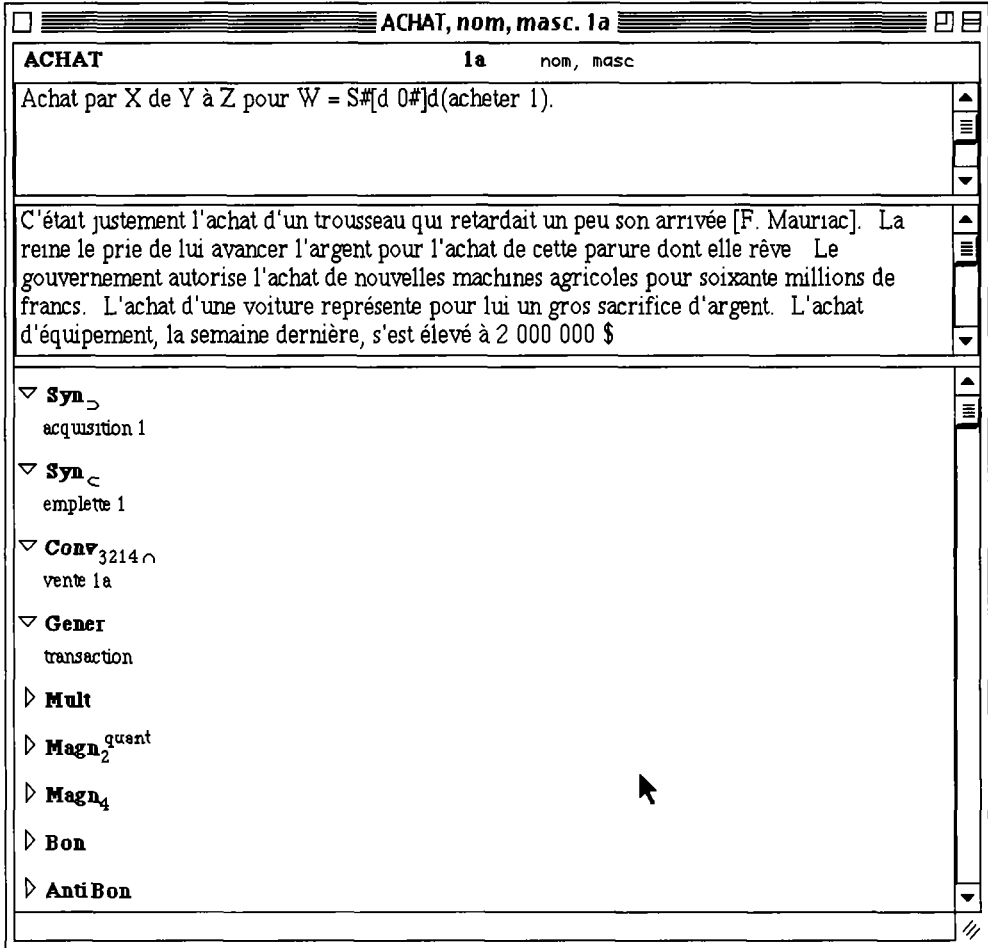


Figure 3 : La fenêtre de lexie pour la lexie achat 1a. En plus des zones de définition et d'exemple, on remarque la zone des fonctions lexicales

RÉCUPÉRATION DES DONNÉES EXISTANTES

En plus de l'éditeur DECID, nous avons développé un outil de récupération des données publiées du DEC. Cet outil part d'un fichier en RTF (Rich Text Format) généré à partir des fichiers Word™ qui ont été utilisés à l'origine pour la création du DEC papier.

Cette récupération n'a pu se faire que semi-automatiquement. Les fichiers en cours de récupération devant être corrigés pour être récupérables. Fort heureusement, les fichiers avaient, dès le départ été créés en utilisant des styles cohérents pour les différents paragraphes décrivant une entrée (définition, régime, etc.). Sans cela, la récupération n'aurait pu avoir lieu.

Certaines difficultés sont dues à l'outil Word™ utilisé. L'absence totale de documentation du format RTF nous a obligé à produire des outils ad hoc. De plus, pour des raisons encore assez obscures, des fichiers d'apparence identiques ont des descriptions RTF différentes. Ainsi, deux paragraphes successifs ayant le même style peuvent apparaître soit comme deux paragraphes indépendants (l'information de style est donnée pour chaque paragraphe) ou comme deux paragraphes identiques (la définition de style n'est donnée qu'au début du premier).

D'autres difficultés sont dues au DEC lui même. Le DEC a été conçu au départ sous une forme papier utilisable par un homme. Aussi, le DEC a été conçu avant tout par sa **présentation**. Aussi, une très grande importance a été accordée à la forme plus qu'à la structure. Ainsi, certaines erreurs dans les documents Word n'étaient pas détectées car elles n'étaient pas visibles sur papier. Par exemple, la définition et la liste d'exemples ont une même forme, mais sont représentées par deux styles différents. Néanmoins, on trouve souvent des erreurs dues à l'identité de forme de ces deux types d'information, l'humain pouvant très facilement faire la différence par le contexte.

Dans les tableaux de régimes, la présentation était contrôlée entièrement par le lexicographe. Ainsi, certaines lignes pouvaient être séparées par une marque de fin de paragraphe, un saut à la ligne, ou même par une succession de tabulations. Certaines valeurs pouvaient tenir sur plusieurs lignes. Dans ce cas, les valeurs apparaissent, dans le fichier séquentiel RTF, de manière entrelacée.

Ces erreurs de présentation et d'édition ont été réglées. Elles ont été fort heureusement assez mineures.

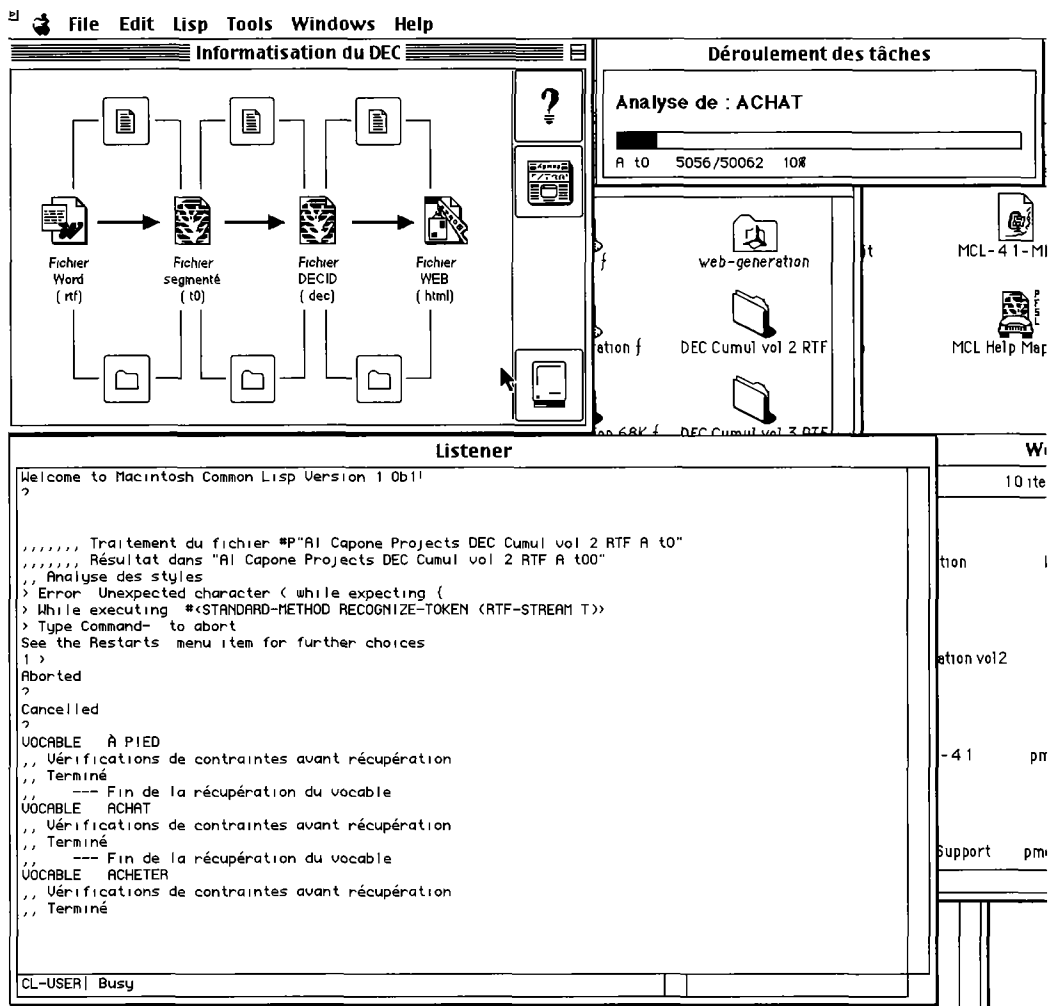


Figure 4 : La plate-forme de récupération des données existantes. La fenêtre en haut à gauche permet au récupérateur de déclencher les traitements. La fenêtre du bas donne des informations sur le traitement en cours

En utilisant la plate-forme de récupération (figure 4), le lexicographe déclenche la récupération d'un fichier RTF ou de tous les fichiers RTF présents dans un dossier. Ces fichiers sont analysés puis, pour chaque vocable détecté, un fichier est créé qui contient les informations sous forme structurée. Pour chaque vocable, le lexicographe peut demander un diagnostic qui lui sera utile pour savoir si les données ont été récupérées de manières satisfaisantes ou si une erreur s'est glissée dans la récupération sans que le processus ne se soit interrompu (figure 5).

L'outil de récupération aide le correcteur en donnant un diagnostic du vocable récupéré. Ainsi, par comparaison avec la version papier, il est aisé de voir ce qui n'a pas fonctionné et pourquoi.

```

;; Vocable      : ACHAT
;; Catégorie   : nom, masc.
;; Possède un tableau résumé : 3 résumés.
;; Pas de note.
;; Lexie      : 1a.
;; Pas de connotations.
;; Possède des informations de régime :
;; Tableau à 4 colonne(s)
;; 1 = X | 2 = Y | 3 = Z | 4 = W
;; 3 | 1 | 2 | 2
;; 3 restriction(s) numérotées.
;; 6 exemples de réalisations
;; Possède des fonctions lexicales :
;; • 11 fonction(s) lexicale(s).
;; Possède des exemples.
;; Lexie      : 1b.
;; Pas de connotations.
;; Possède des informations de régime :
;; Tableau à 3 colonne(s)

;; 1 = X | 2 = Y | 3 = Z
;; 3 | 2 | 1
;; 2 restriction(s) numérotées.
;; 2 exemples de réalisations
;; Possède des fonctions lexicales :
;; • 14 fonction(s) lexicale(s).
;; Possède des exemples.
;; Lexie      : 2.
;; Pas de connotations.
;; Possède des informations de régime :
;; Tableau à 1 colonne(s)
;; 2 = X
;; 2
;; 0 restriction(s) numérotées.
;; 1 exemples de réalisations
;; Possède des fonctions lexicales :
;; • 3 fonction(s) lexicale(s).
;; Possède des exemples.

```

Figure 5 : Le diagnostic de récupération d'un vocable

EXPLOITATION DES DONNÉES INFORMATISÉES

Les données ainsi informatisées ont été exportées sous forme HTML. Nous avons ainsi pu produire automatiquement un site Web complet présentant le DEC sous une **présentation** analogue à celle utilisée dans la version papier. Les figures 6 et 7 représentent une page du DEC vue par un navigateur standard.

Cette version HTML du DEC est, comme le DEC au format Word, destinée à un usage humain. Nous avons adopté une forme aussi proche que possible de la forme originale. Mais cette présentation pose différents problèmes. En effet, le format HTML ne permet pas, de manière simple, de préciser effectivement une forme. C'est le navigateur qui, en dernier ressort, effectue la présentation. Cela pose différents problèmes :

- les tables, n'apparaissent pas de la même manière suivant les navigateurs utilisés. Cela peut mener à des colonnes trop larges ou trop étroites;
- tous les navigateurs ne savent pas forcément interpréter et présenter les informations en indice ou en exposant.

Enfin, indépendamment de la compatibilité des différents navigateurs, la version HTML du DEC pose des problèmes intrinsèques. Ainsi, certains caractères sont propres au DEC (ex : k et l qui délimitent les locutions ou ' et " qui délimitent les sémantèmes). Ces caractères ne sont présents dans aucune fonte standard. Actuellement, HTML ne permet pas d'inclure une fonte ou une description de caractère qui soit portable. On peut

indiquer au navigateur d'utiliser une fonte particulière, mais celle-ci doit être présente dans le système du client. L'utilisation d'une image pour ces caractères peut être envisagée, mais le client n'aura pas une bonne présentation s'il décide de changer la taille des caractères affichés (l'image ne grandira pas en fonction).

The screenshot shows a Netscape browser window titled "Netscape: ACHAT nom, masc.". The address bar contains the file path: "file:///Grincieux/Projects/web%20Generation/R%8Ecup-Dec%20final/DEC%20II-final/DEC%202-htm". The main content area displays the word "ACHAT, nom, masc." followed by three numbered definitions:

- 1a. S_0 (acheter 1) [*J'achète par Marie d'une robe*]
- 1b. Activité commerciale - ensemble de tous les achats 1a ... [*Les achats, par l'EPSS, de céréales au Canada*]
2. Ce que X a acheté 1 ... [*Pierre m'a montré ses derniers achats*]

Below the definitions, it states: "1a. Achat par X de Y à Z pour W = S_0 (acheter 1)".

The word "Régime" is centered, followed by a table with four columns labeled "1 = X", "2 = Y", "3 = Z", and "4 = W".

1 = X	2 = Y	3 = Z	4 = W
1. de N	1. de N	1 à N	1. de Num
2. par N		2. Loc in N	N
3 A _{poss}			2. pour N

Below the table, there are several entries on the left and their corresponding explanations on the right:

- 1) $C_{3,2}$ N désigne un commerçant ou une entreprise commerciale
- 2) $C_{4,2}$ sans C_2 impossible
- 3) $C_{1,2}$ sans C_2 non souhaitable
- C_1 *les achats de Pierre, ses achats*
- C_2 *un achat de marchandises*
- $C_3 + C_4$ *un achat de 15 dollars à cette entreprise <chez un épicier>*
- $C_1 + C_2 + C_3 + C_4$ *l'achat par Marie d'une robe chez la couturière pour cinquante dollars*
- Impossible **l'achat pour 3 000 dollars (2) (= l'achat de 3 000 dollars / l'achat d'une fourneuse pour 3 000 dollars)*
- Non souhaitable ** les derniers achats par l'entreprise (3) (= les derniers achats de l'entreprise / les derniers achats d'équipement par l'entreprise)*

Figure 6 : Le vocable ACHAT nom, masc vu par un navigateur standard

The screenshot shows a Netscape browser window with the title "Netscape: ACHAT nom, masc.". The address bar shows a file path. The main content area is titled "Fonctions lexicales" and contains the following table:

Syn ₃	acquisition 1
Syn ₂	emplette 1
Conv ₃₂₁₄ _n	vente 1a
Gener	transaction
Mult	achats 1b
Magn ₂ _{quant}	massif
Magn ₄	important, grand prépos
Bon	bon
AntiBon	mauvais
Oper ₁	faire, effectuer [ART ~]
Func ₄	s'élever [à N]

Below the table is a section titled "Exemples" with the following text:

C'était justement l'achat d'un trousseau qui retardait un peu son arrivée [F. Mauriac]. La reine le prie de lui avancer l'argent pour l'achat de cette perure dont elle rêve. Le gouvernement autorise l'achat de nouvelles machines agricoles pour soixante millions de francs. L'achat d'une voiture représente pour lui un gros sacrifice d'argent. L'achat d'équipement, la semaine dernière, s'est élevé à 2 000 000 \$.

Figure 7 : Les fonctions lexicales de la lexie ACHAT nom, masc 1a

CONCLUSION

À l'occasion de l'action de recherche partagée NADIA-DEC, nous avons donc pu informatiser le *Dictionnaire explicatif et combinatoire du français contemporain*. Pour cela, nous avons récupéré semi automatiquement la totalité des entrées des volumes II et III du DEC.

Nous avons aussi créé un outil d'édition spécialisé pour le DEC. Cet outil offre des avantages certains aux lexicographes, mais il contraint trop la structure du dictionnaire et ne pourra être utilisé tant que l'édition du DEC se fera dans l'optique d'une recherche en lexicologie (structure en cours de définition).

Néanmoins, les travaux effectués seront très utiles pour un passage en phase de production du DEC ou d'un dictionnaire dérivé.

Ce travail a constitué une étape importante dans nos recherches sur des outils pour lexicographes. Elle nous a permis de mettre en œuvre nos méthodes sur un dictionnaire très complexe. Nous avons ainsi pu valider certains choix. Néanmoins, nous avons pu

voir que la construction d'outils spécialisés fige la structure informatique utilisée. Or, dès que l'on travaille avec des dictionnaires assez compliqués, la possibilité de remise en cause des structures informatiques en cours d'édition d'un dictionnaire est nécessaire.

Aussi, nous souhaitons orienter nos recherches sur des méthodes génériques de création d'outils spécialisés pour lexicographes. Cette généricité nous permettra d'offrir des outils évolutifs et rendra plus facile les recherches de lexicologie.

RÉFÉRENCES

- MEL'CUK, Igor, CLAS, André et Alain POLGUÈRE (1995) : *Introduction à la lexicologie explicative et combinatoire*, coll. «Universités francophones» et «champs linguistiques», Louvain-la-Neuve, AUELF-UREF et Duculot.
- SÉRASSET, Gilles (1994) : *SUBLIM : un système universel de bases lexicales multilingues et NADIA : sa spécialisation aux bases lexicales interlingues par acceptions*, Thèse nouveau doctorat, Université Joseph Fourier-Grenoble 1, 194 p.
- SÉRASSET, Gilles (1995) : «Informatisation du *Dictionnaire explicatif et combinatoire* : le projet NADIA-DEC», *Lexicomatique et dictionnaires*, Actes des IV^{es} Journées scientifiques du réseau LTT, Lyon, 28-30 septembre 1995, pp. 205-215.
- SÉRASSET, Gilles (1996) : «Un éditeur pour le Dictionnaire explicatif et combinatoire du français contemporain», *Journées lexique du PRC-CHM*, Grenoble, 13-14 novembre 1996, pp. 131-138.
- SÉRASSET, Gilles (1997) : «Informatisation du Dictionnaire explicatif et combinatoire», *TALN'97*, Grenoble, 12-13 juin 1997, pp. 194-198.
- SÉRASSET, Gilles et Étienne BLANC (1993) : «Une approche par acceptions pour les bases lexicales multilingues», *TA-TAO : recherches de pointe et applications immédiates*, Actes des III^{es} Journées scientifiques du réseau LTT, Montréal, 30 septembre-2 octobre 1993, pp. 65-84.
- SÉRASSET, Gilles et Alain POLGUÈRE (1997) : «Outils pour lexicographes : application à la lexicographie explicative et combinatoire», *RIAO'97*, Montréal, 25-27 juin 1997, pp. 701-708.

TAO ET THÉORIES LINGUISTIQUES : INSTITUTIONS GRAMMATICALES*

Philippe BLACHE et Jean-Yves MORIN

LPL - CNRS, Aix-en-Provence, France et Université de Montréal, Canada

INTRODUCTION

Si l'on présuppose

- (a) qu'il est inévitable en TALN (et plus spécifiquement en TAO) de tenir compte des propriétés fondamentales des langues naturelles (paradoxalement, ce ne semble pas être un truisme pour l'ensemble des chercheurs dans le domaine),

si, de plus, l'on admet que

- (b) les théories et les descriptions linguistiques actuelles arrivent à identifier certaines de ces propriétés,

si, enfin, l'on croit que

- (c) ces différentes théories et descriptions ne sont pas nécessairement contradictoires, mais présentent différents *points de vue* sur la réalité,

on peut vouloir intégrer ces différentes descriptions, les rendre compatibles l'une à l'autre et les exploiter dans un système efficace et transparent.

Ce que nous nous proposons de construire, c'est un cadre général basé sur les notions de *dimensions*, d'*objets* et de *contraintes* permettant de réaliser à la fois l'œcuménisme théorique nécessaire et l'intégration des différents niveaux de représentation. On tentera donc de développer ce que nous appellerons des INSTITUTIONS GRAMMATICALES, qui fourniront de tels cadres généraux permettant de caractériser et de comparer la structure logique et calculatoire de ces théories et descriptions linguistiques.

* Une version préliminaire de ce travail a été présentée aux V^{es} Journées scientifiques du réseau LTT *La mémoire des mots* à Tunis, en septembre 1997. Ce travail a bénéficié de l'aide du réseau LTT à l'équipe «TAO et théories linguistiques» (Anne Abeillé, Philippe Blache, Jean-Yves Morin et Éric Wehrli) ainsi que d'une subvention de recherche du fonds FCAR du Gouvernement du Québec.

Il y a une quinzaine d'années, Goguen et Burstall (1984) introduisaient en logique informatique la notion d'*institution*.

«This paper shows how some parts of computer science can be done in any suitable logical system, by introducing the notion of an **institution** as a precise generalization of the informal notion of a "logical system". A first main result shows that if an institution is such that interface declarations expressed in it can be glued together, then **theories** (which are just sets of sentences) in that institution can also be glued together. A second main result gives conditions under which a theorem prover for one institution can be validly used on theories from another; this uses the notion of an institution morphism. A third main result shows that institutions admitting free models can be extended to institutions whose theories may include, in addition to the original sentences, various kinds of constraints upon interpretations; such constraints are useful for defining abstract data types, and include so-called "data", "hierarchy", and "generating" constraints. Further results show how to define institutions (sic) that mix sentences from one institution with constraints from another, and even mix sentences and (various kinds of) constraints from several different institutions.»

[...]

«Informally, an institution consists of

- a collection of signatures (which are vocabularies for use in constructing sentences in a logical system) and signature morphisms, together with for each signature Σ ,
- a set of Σ -sentences,
- a set of Σ -models, and
- a Σ -satisfaction relation, of Σ -sentences by Σ -models

such that when you change signatures (with a signature morphism), the satisfaction relation between sentences and models changes consistently.»

(Goguen & Burstall 1984 : 221-222, l'emphase est dans l'original.)

En termes très simples, cela équivaut à dire que si l'on sait de quoi deux institutions A et B parlent (leurs signatures¹), et que l'on sait traduire d'une institution à l'autre, on peut

¹ Rappelons qu'une signature est un couple $\text{Sig} = \langle S, \Sigma \rangle$ où

S est un ensemble (généralement structuré) de *sortes* ou *types* et

Σ est une famille de fonctions indexée par $S^* \times S$.

Pour $\sigma \in \Sigma_{as}$, on note

$\sigma : a \rightarrow s$

a est l'*arité* de σ (le nombre d'arguments de σ et leur type respectif)

s est la *sorte*

la paire $\langle a, s \rangle$ est le *rang* associé à σ .

En termes informels, une signature distribue les termes d'une théorie sur des ensembles ordonnés d'arguments typés (dans cette formulation, tous les termes sont des fonctions, les constantes sont des fonctions sans arguments, les prédicats à n arguments sont des fonctions de S^n dans V (un ensemble de valeurs de vérité). En fait, l'objectif de ce type de définition formelle (pour des lexiques abstraits) est tout à fait analogue à ce que font les définitions de type DEC (pour des lexiques concrets) qui attribuent à chaque lexie un type et une liste ordonnée d'arguments typés par des étiquettes sémantiques (cf. Mel'cuk et al. 1995; Miličević 1997). Par exemple, pour $\text{VENDRE}_{1,1}$ dans la syntaxe

combiner de façon cohérente les énoncés de A et ceux de B et déterminer des *conditions de collage* entre ces énoncés.

Pratiquement, il s'agit d'identifier les *prédicats théoriques* utilisés dans ces théories/descriptions, d'étudier leur sémantique descriptive (i.e., les DOMAINES d'entités qu'ils définissent/décrivent), leurs propriétés calculatoires (i.e., COMPLEXITÉ théorique et effective) et leur sémantique opératoire (représentabilité, directe ou indirecte, dans des implantations effectives) et d'élaborer des moyens d'intégrer les représentations correspondantes dans des environnements effectivement utilisables.

QUELQUES PROBLÈMES

Le volume et la disparité des connaissances linguistiques nécessaires en TAO pose de délicats problèmes d'ÉLABORATION et d'INTÉGRATION. En ce qui concerne l'ÉLABORATION, la plupart des travaux linguistiques menés depuis les trente dernières années portent plutôt sur l'interprétation théorique de l'analyse de phénomènes particuliers dans diverses langues que sur la construction de descriptions approfondies et couvrantes pour une langue en particulier². Ceci tient d'une part à la nature extrêmement complexe des phénomènes linguistiques, qui ne se prêtent pas à une analyse globale rapide, ou même incrémentale, sauf dans certains domaines spécifiques (phonologie, morphologie, une partie du lexique) et d'autre part à des facteurs socio-historiques de développement de la discipline qui favorisent les travaux théoriques plutôt que descriptifs. On ne dispose donc pas d'études descriptives approfondies qui aient, par exemple, la couverture empirique de l'*Essai de grammaire de la langue française* de Damourette et Pichon (1911-1940). Ce travail d'élaboration de descriptions fines, approfondies et couvrantes est donc prioritaire (Abeillé, Godard & Miller, à paraître). Le problème d'INTÉGRATION, quant à lui, présente deux facettes. D'une part, les connaissances linguistiques sont partitionnées en *niveaux*, plus ou moins autonomes qu'il est essentiel de pouvoir interfacer les uns aux autres de façon efficace. D'autre part, même à l'intérieur d'un niveau, les *représentations* utilisées peuvent être très différentes d'une perspective théorique à l'autre et les niveaux eux-mêmes peuvent varier. Ainsi, pour caractériser l'*ordre des constituants* à l'intérieur d'un syntagme, une théorie comme la théorie principes-paramètres classique (Chomsky 1981, PP) utilisera essentiellement trois paramètres binaires (tête initiale/finale, assignation de Cas vers la gauche/droite et assignation de θ -rôle vers la gauche/droite)³,

simplifiée du DiCo (Mel'cuk et al. 1995 : 224) : action sociale : personne X ~ Y à personne Z pour argent W.

Les termes *action sociale*, *personne* et *argent* sont des types contraignant respectivement le prédicat lui-même (*action sociale*), le premier et le troisième argument (*personne X et personne Z*) ainsi que le quatrième argument (*argent W*).

² Les travaux des différentes équipes travaillant sur des *lexiques-grammaires* (Gross, 1975) constituent une exception d'autant plus notoire qu'elle semble isolée.

³ Cf. Travis (1989) et Fodor & Crain (1990). Les paramètres en question ont des effets ailleurs que dans le domaine des contraintes d'ordonnement. Ceci constitue une *qualité* pour les tenants de la théorie PP, puisque cela ajoute à la «richesse de sa structure déductive». Du point de vue de la représentation des connaissances linguistiques, c'est plutôt un défaut, puisque cette propriété empêche de maintenir la transparence fonctionnelle des descriptions et donc la modularité des représentations. Ce(tte) défaut/qualité est maintenu(e) dans le programme minimaliste (Chomsky 1995; Epstein 1996 éd., PM).

tandis que des théories basées sur l'unification⁴, comme la grammaire syntagmatique généralisée (Gazdar et al. 1985, GSG) ou la grammaire syntagmatique endocentrique (Pollard & Sag 1994, GSE) utilisent un type particulier de règles/schémas, les PL-règles/schémas.

En français et en japonais, les sujets précèdent en général le syntagme verbal central de la proposition, alors qu'en malgache les sujets suivent le syntagme (verbal ou non) central de la proposition. En français et en malgache, les compléments non pronominaux apparaissent après la tête lexicale (verbe, nom, adjectif ou préposition) dont ils dépendent, alors qu'en japonais, ils apparaissent avant⁵.

- (1) (a) Je suis allé à la montagne avec mon frère hier.
- (b) [P [P [SN *je* SN] [SV [V *suis allé* V] [SP [Prép à Prép] [SN [Dét *la* Dét] [N *montagne* N] SN] SP] [SP [Prép *avec* Prép] [SN [Dét *mon* Dét] [N *frère* N] SN] SP] SV] P] [SP_{Adv} [Adv *hier* Adv] SP_{Adv}] P]
- (2) (a) Nandeha tany an-tendrombohitra niaraka tamin'ny rahalahiko aho omaly.
N_{PASSÉ}-aller t_{PASSÉ}-LOCLOC-montagne t_{PASSÉ}-être ensemble t_{PASSÉ}-avec Dét frère-POSS_{Ips} PRO_{Ips} hier
- (b) [P [P [SV [SV [V *nandeha* V] [SP [Prép *tany* Prép] [SN [N an-tendrombohitra N] SN] SP] SV] [SV [V *niaraka* V] [SP [Prép *amin* Prép] [SN [Dét *ny* Dét] [N *rahalahiko* N] SN] SP] SV] SV] [SN *aho* SN] P] [SP_{Adv} [Adv *omaly* SP_{Adv}] P]
- (3) (a) Kinô boku wa nii-san to yama ni nobori-mashi-ta.
Hier moi TOP frère-HON COM montagne DAT monter-AUX-PASSÉ
- (b) [P [SP_{Adv} [Adv *Kinô* Adv] SP_{Adv}] P [SN [N *Boku* N] [Postp *wa* Postp] SN] [SV [SN [N *nii-san* N] [Postp *to* Postp] SN] [SN [N *yama* N] [Postp *ni* Postp] SN] [V *noborimashita* V] SV] P] P]

⁴ Cf. Abeillé (1993).

⁵ Pour des raisons de simplicité, nous avons choisi une représentation par crochets étiquetés indentés de la structure syntagmatique. Cette représentation utilise des étiquettes qui sont des abréviations des catégories utilisées en GSG/GSE plutôt que de celles utilisées en PP. Une représentation de la structure syntagmatique dans le cadre PP serait nettement plus complexe que celle ci-dessus, mettant en jeu plusieurs niveaux (D-structure, S-structure, etc.), donc plusieurs représentations distinctes, chacun de ces niveaux contiendrait une multitude de projections fonctionnelles (SC, SI, SD) avec une grande quantité de catégories vides (INFL, COMP, DÉT, etc.), tout en n'étant pas nécessairement plus informative. Pour ce qui nous concerne ici (la distribution des catégories lexicales effectivement réalisées relativement aux catégories syntagmatiques), une représentation de S-structure dans le cadre PP, dépouillée de tous ces artifices (i.e., projections fonctionnelles, catégories vides et traces de mouvements) serait essentiellement congruente à la représentation donnée ici

Dans le cadre PP, on pourrait poser

- (a) que le français et le japonais ont un sujet extrait en SS, alors que le malgache aurait un sujet extrait en FL et
 - (b) que le français et le malgache sont des langues à tête initiale, avec assignation de Cas et de θ -rôle vers la droite, alors que le japonais est à tête finale, avec assignation de Cas et de θ -rôle vers la gauche⁶.
- (4) Paramètres PP

(a) Français

SUJET	TÊTE	CAS	Θ -RÔLE
SS	Initiale	Droite	Droite

(b) Malgache

SUJET	TÊTE	CAS	Θ -RÔLE
FL	Initiale	Droite	Droite

(c) Japonais

SUJET	TÊTE	CAS	Θ -RÔLE
SS	Finale	Gauche	Gauche

Dans le cadre GSG/GSE, on pourrait poser que le SUJET est initial en français et en japonais et final en malgache et que ce sont les CATÉGORIES LEXICALES (têtes, mais également spécificateurs — déterminants, auxiliaires, modificateurs, etc.—) qui doivent apparaître avant les catégories syntagmatiques en français et en malgache, alors qu'elles apparaissent après ces dernières en japonais.

⁶ Il ne s'agit pas ici d'une analyse mais simplement d'une illustration. En fait, la description des contraintes d'ordre des mots en français, en malgache et en japonais dans le cadre PP serait beaucoup plus complexe que ne le laissent soupçonner ces tableaux. L'extraction devrait être contrainte de façon explicite. La valeur des quatre paramètres serait également paramétrisable. En français, par exemple, l'assignation du Cas accusatif par V ou oblique par Prép doit se faire vers la droite, mais l'assignation du Cas nominatif par INFL, vers la gauche. De plus, certaines têtes devraient être initiales, d'autres finales. On notera également que ce système rend nécessaire une multiplication de niveaux, puisqu'il exclut a priori des structures «plates» où la tête est enrobée de part et d'autres de constituants (spécificateurs d'un côté et compléments de l'autre, par exemple [SN Dét N SP] ou [SV Aux V SN] en français). Une telle structure minimale ne serait disponible que pour une langue où le paramètre de tête (initiale/finale) ne serait pas actif.

(5) (a) Français, japonais

PL : [SUJET(α)] < α

(b) Malgache

PL : α < [SUJET(α)]

(6) (a) Français, malgache

PL : [NIVEAU : **lexical**] < \neg [NIVEAU : **lexical**]

(b) Japonais

PL : \neg [NIVEAU : **lexical**] < [NIVEAU : **lexical**]

Si l'on tentait de calquer en partie la théorie PP en GSG/GSE, on pourrait postuler un trait [TÊTE : α] caractérisant les têtes de syntagmes⁷ et avoir des PL-règles comme en (7).

(7) (a) Français

PL : [TÊTE : α] < \neg [TÊTE : α]

(b) Japonais

PL : \neg [TÊTE : α] < [TÊTE : α]

⁷ En fait, Zwicky (1988) postule quelque chose de semblable. Dans le système de Gazdar et al. (1985), ce trait serait, pour les têtes lexicales, une abréviation de [SOUS-CAT : n], où n est l'index d'une règle. Quant aux têtes non lexicales, si elles peuvent porter le trait [SOUS-CAT : n], comme le proposent Pollard et Sag (1987), elles sont identifiables de façon analogue. Sinon, il faudrait construire un trait ad hoc. Cependant, on remarque que leur comportement est différent de celui des têtes lexicales. Par exemple, les têtes lexicales précèdent leurs compléments, mais SV suit son sujet en français. C'est d'ailleurs ce qui nous amène à exprimer les contraintes d'ordre en termes de NIVEAU (lexical ou non lexical) plutôt que de fonction (tête vs spécificateurs ou compléments).

Si, par exemple, on cherche à intégrer une description du SN (français ou japonais) suivant le cadre GSG/GSE et une description de la proposition conforme au cadre PP, il faudra d'abord opérer les coercitions nécessaires dans un sens (PARAMÈTRESTÊTE/CAS/Θ > PL-RÈGLES) ou dans l'autre (PL-RÈGLES > PARAMÈTRESTÊTE/CAS/Θ). Une telle opération ne pose pas de problème de principe, mais exige un investissement considérable. On pourrait penser que le problème de l'intégration des REPRÉSENTATIONS peut être évité en choisissant une fois pour toutes une théorie et en s'y tenant⁸. L'expérience montre que tel n'est pas le cas⁹. La description de fragments importants de plusieurs langues naturelles présuppose la possibilité de réinterpréter les représentations de diverses théories dans la théorie choisie. De plus, même à l'intérieur d'une théorie donnée, la dynamique de la description entraîne des changements non négligeables de la théorie sous-jacente. Pourquoi dès lors ne pas faciliter directement cette réinterprétation? C'est précisément ce qu'une analyse en termes de DOMAINES, comme celle esquissée ci-haut pour le domaine des contraintes d'ordre permet de faire. Les institutions grammaticales sont les recueils de ces différents domaines, qui peuvent aussi bien être définis a priori qu'enrichis par l'ajout de nouvelles descriptions. Ainsi, pour la description syntaxique générale, on posera les domaines suivants :

⁸ Quant au problème de l'intégration des NIVEAUX, il demeure, de toute façon, incontournable.

⁹ Jusqu'à tout récemment, le groupe EUROTRA semblait favoriser un tel type d'approche. Cf. Arnold & des Tombe (1987).

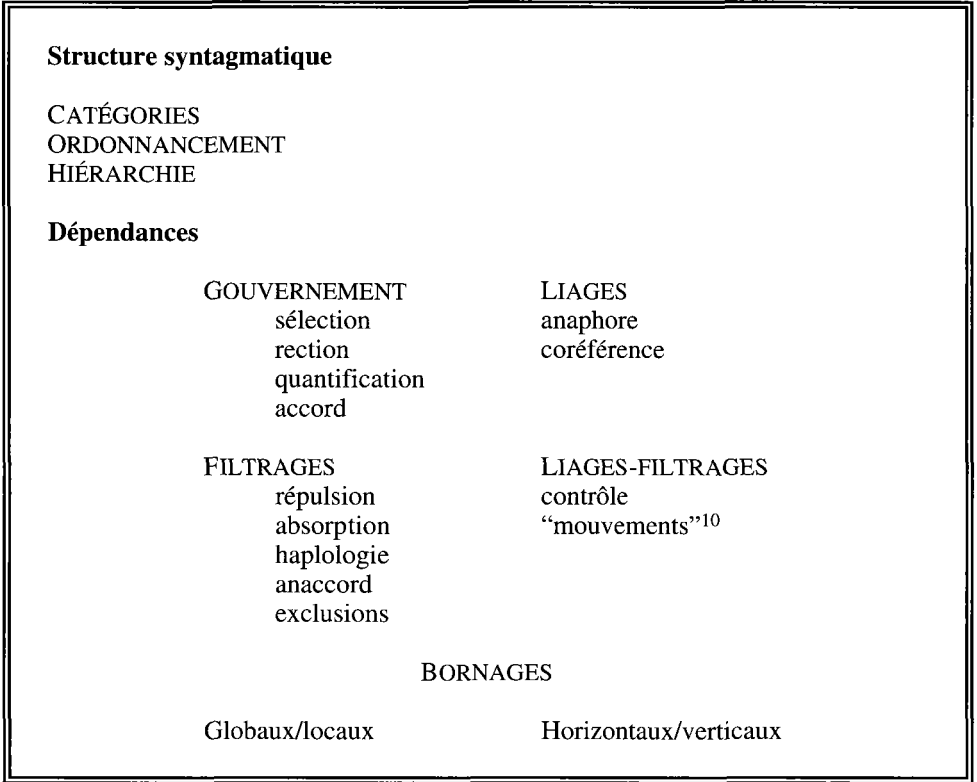


Figure 1 : Domaines syntaxiques

Cette organisation est très souple. Si, par exemple, on voulait introduire la notion de *rayon*¹¹ de Damourette et Pichon dans une description formalisée, elle trouverait tout de suite sa place distribuée dans les domaines des CATÉGORIES et de la rection et l'on pourrait immédiatement voir qu'elle correspond aux valeurs du trait TFORME, telles qu'elles apparaissent comme valeurs de SOUS-CAT dans GSE. Si même l'on voulait introduire une notion ne correspondant à aucun des domaines prédéfinis, comme la

¹⁰ On regroupe sous ce terme inspiré de la métaphore transformationnelle, les liages-filtrages caractérisés en PP/PM comme A, \bar{A} ou parasites. En GSG/GSE, ou GLF il s'agit non pas de liages entre position pleine et vide, mais entre structures informationnelles, qui peuvent ou non correspondre à un noeud spécifique d'un arbre syntagmatique.

¹¹ RAYON : classification des compléments selon leur introduction : «Je vois *mon père*» — direct; «Il a couru *vers la porte*» — vers; «Priez *pour moi*» — pour; «Arriver à *la ville*» — à; «Le curé de *Bazeille*» — de; «Je sors *le matin*» hors —.

Damourette, J. et E. Pichon (1991-1940) : *Des mots à la pensée. Essai de grammaire de la langue française. Compléments*, 13, D'Artrey, Paris, 1971.

(Le trait — doit être lu *rayon*. J'ai corrigé certaines marques typographiques incohérentes. JYM)

*visée*¹², toujours empruntée à Damourette et Pichon, on pourrait créer un nouveau type de liage correspondant et s'apercevoir que ce liage correspond à la notion de *première paire d'éléments dans la liste d'oblicité* ARGS en GSE.

On remarquera que ces domaines syntaxiques n'incluent pas de domaines correspondant directement aux notions configurationnelles, si importantes en PP/PM. On pourrait soit introduire ces notions configurationnelles comme portant sur de nouveaux sous-domaines des BORNAGES¹³, soit les faire dériver entièrement d'autres notions non configurationnelles. Comme les notions configurationnelles sont omniprésentes dans toutes les versions de PP/PM¹⁴, elles affectent à peu près tous les domaines (même l'information grammaticale morphologique y est codée configurationnellement). Il semble donc difficile de la factoriser. De plus, le rôle filtrant de ces notions est extrêmement faible : dans tout arbre non dégénéré (i.e., tout arbre branchant) *tous* les noeuds, sauf la racine, sont dans la relation de c-commande, soit comme c-commandeur, soit comme c-commandé. Aussi avons-nous opté plutôt pour la réduction systématique à des notions non configurationnelles (Morin 1996).

REPRÉSENTATION DES CONNAISSANCES LINGUISTIQUES : GRAMMAIRES D'OBJETS ET DE CONTRAINTES

Classiquement, il existe deux grands types de représentation des connaissances, qu'il s'agisse de connaissances linguistiques ou non : les REPRÉSENTATIONS PAR OBJETS et les REPRÉSENTATIONS PAR CONTRAINTES. Les REPRÉSENTATIONS PAR OBJETS définissent des univers d'objets, dotés de propriétés, entrant dans des relations et, éventuellement, capables d'actions (acteurs). Ces représentations mettent l'accent sur l'autonomie des objets, sur leur regroupement en classes organisées hiérarchiquement. Les objets d'une hiérarchie de classes partagent les uns avec les autres des propriétés héritées de leurs classes ancestrales communes. Ils peuvent prendre des valeurs par défaut ou déléguer à leurs ancêtres différentes fonctions. Cette conception semble intuitivement correspondre plus directement à notre ontologie naïve. Les REPRÉSENTATIONS PAR CONTRAINTES définissent des univers de relations abstraites ou contraintes. Les individus ou objets n'y existent et n'y sont identifiables qu'à travers les contraintes qu'ils satisfont. Ce type de représentation favorise l'abstraction et l'inférence. Cette conception semble correspondre à une ontologie plus formelle. Au niveau linguistique, on peut dire qu'un DICTIONNAIRE constitue une REPRÉSENTATION PAR OBJETS du langage, alors qu'une GRAMMAIRE en constitue plutôt une REPRÉSENTATION PAR CONTRAINTES.

¹² VISÉE : rapport établi par un verbe entre son sujet et son about, c'est-à-dire celui de ses compléments auquel il tend par la force de sa signification (objet, attribut, etc.). Damourette, J. et E. Pichon (1991-1940) : *Des mots à la pensée. Essai de grammaire de la langue française. Compléments*, 13, D'Artrey, Paris, 1971.

¹³ Les contraintes configurationnelles, comme *c-commande*, *gouvernement*, etc., ont un rôle essentiellement filtrant et non pas constructif, comme on l'a montré dans Morin (1996).

¹⁴ Sauf peut-être la réduction de *c-commande* à des notions transdérivationnelles d'Epstein (1996, réd.). Mais cette dernière introduit tellement de complexité cachée qu'il s'agit purement d'un exercice académique.

Objets

Représentation à base d'objets : grammaires à base lexicale

Plusieurs théories grammaticales récentes¹⁵ partent de l'hypothèse que les connaissances grammaticales (du moins celles qui sont spécifiques à une langue) sont essentiellement rassemblées dans le lexique. Cependant, outre un certain nombre de remarques et de travaux intéressants, mais de nature plutôt programmatique, sur le rôle et le fonctionnement du lexique, sur sa position relative face aux autres composantes, sur sa structure générale ou sur la forme des entrées lexicales, très peu de chercheurs se sont penchés sur la construction effective de lexiques explicites¹⁶. Dans les théories linguistiques basées sur la notion d'*information grammaticale*, le rôle du lexique est particulièrement important, puisque celui-ci y constitue le répertoire principal (tant par l'envergure que par la complexité) de l'information grammaticale spécifique à une langue. De plus, dès que l'on se préoccupe de la *couverture* d'une description, tout autant que de sa rigueur, les problèmes associés à la construction de vastes lexiques deviennent incontournables¹⁷.

En traitement des langues naturelles, d'autres facteurs encore contribuent à donner au lexique un rôle primordial. Le lexique contient l'ensemble des objets linguistiques directement accessibles et donc de coût constant, en termes de ressources. Les autres objets (structures, relations) ne sont accessibles que par construction¹⁸ et sont donc de coût variable (éventuellement prohibitif) toujours en termes de ressources. En gros, il est plus facile de vérifier une information que de la (re)construire. Ce fait est reconnu, au moins implicitement, par la plupart des chercheurs du domaine.

À un niveau abstrait, un lexique constitue une *représentation par objets* des connaissances grammaticales. C'est un ensemble d'objets autonomes appartenant à des classes organisées en des réseaux plus ou moins serrés (e.g. catégories, catégories lexicales, catégories lexicales majeures ou verbes, verbes transitifs, verbes transitifs à double objet, etc.)¹⁹. Les objets des sous-classes héritent des propriétés de leurs superclasses (ou leur délèguent des propriétés). L'information sur le comportement syntaxique d'un objet (par exemple, la transitivité d'un verbe) est partiellement stockée dans la classe de l'objet et n'a pas à être répétée complètement dans l'entrée lexicale de celui-ci. Si un verbe appartient à la classe des verbes transitifs, il aura un comportement syntaxique de verbe transitif, dont les détails sont fixés une fois pour toutes et n'ont pas à être répétés. Seules les propriétés plus idiosyncratiques auront à être spécifiées dans

¹⁵ Par exemple PP, où lexique et paramètres sont fondamentaux, la grammaire lexicale-fonctionnelle (GLF) ou GSE.

¹⁶ Notons cependant les travaux de Gross (1975) sur les *lexiques-grammaires* et ceux de Mel'cuk (1984, 1988, 1992) sur les *dictionnaires explicatifs et combinatoires* (DEC) qui, malgré leur intérêt, s'intègrent difficilement aux théories syntaxiques issues de la grammaire générative, qui nous intéressent plus particulièrement ici.

¹⁷ Sur le problème de la couverture, cf. Bouchard, Emirikian et Morin (1991).

¹⁸ Si l'on compare une analyse à une démonstration, les entrées lexicales constituent autant d'axiomes, directement démontrés, alors que les structures ne peuvent être démontrées que de façon constructive, à partir des règles du système, ce qui constitue une source de complexité calculatoire.

¹⁹ GSE étend cette structure de réseau en des hiérarchies de TYPES universelles.

l'entrée lexicale même (par exemple, le fait de régir un cas particulier pour l'objet direct). Ce type de représentation permet de construire des descriptions de façon très modulaire, incrémentale et dynamique et d'exprimer des valeurs par défaut. Ainsi, une propriété qui peut sembler idiosyncratique, comme le fait d'avoir une forme médio-passive en *se* sera d'abord notée à l'intérieur des entrées lexicales spécifiques (*vendre, laver*, etc.). Par la suite, elle pourra migrer hors de ces entrées et former une nouvelle classe où seront rassemblées les propriétés générales des médio-passifs, les entrées ne comportant plus qu'une référence à une fonction générique *médio-passif* indiquant que le verbe en question est dans le domaine de cette fonction.

En fait, toute la connaissance grammaticale peut être représentée en termes d'objets. Non seulement les objets lexicaux sont associés à des règles (ou des familles de règles²⁰), mais les règles elles-mêmes peuvent être vues comme des objets abstraits. Les règles et les principes sont organisés en réseaux de classes et de sous-classes (correspondant en partie aux «modules» d'une théorie comme PP ou GSG/GSE). Elles ont des propriétés qui peuvent être héritées (ou déléguées). Ainsi, une règle ou un principe en PP a différents attributs : une strate ou un ensemble de strates d'application (DS, SS, FL, etc.), des paramètres, etc. Par exemple, l'assignation de CAS en PP peut être vue comme une classe d'objets ayant des paramètres de CAS, (eux-mêmes formés d'une VALEUR (par défaut *acc*) et d'un ASSIGNATEUR), d'une DIRECTION et d'une STRATE d'application (qui serait, par défaut la *S-structure*).

(8) Assignation de CAS (PP)

CAS :	[VALEUR : <i>c</i> (<i>acc</i>)]
ASSIGNATEUR :	{	<i>V,</i> <i>Infl,</i> <i>Prép,</i> ...	}
DIRECTION :	{	<i>gauche</i> <i>droite</i>	}
STRATE :		<i>s</i> (<i>S-structure</i>)	

De même, en GSG/GSE, une DI-règle/clause du DI-principe a une IDENTITÉ formée d'un NOM ET d'un INDEX ET d'une FAMILLE, une FONCTION —décomposition, adjonction, coordination—, et un CONTENU, CONSTITUÉ d'une MÈRE (ou PARTIE GAUCHE), d'un ensemble de FILLES (ou PARTIE DROITE), qui peuvent être des FILLES-TÊTES (dont le NIVEAU est, par défaut, lexical), des FILLES-COMPLÉMENTS, des FILLES-SPECIFICATEURS ou des FILLES-ADJOINTES. La présence de tous ces attributs est héritée, pour chaque DI-règle, du fait même qu'il s'agisse d'une DI-règle.

²⁰ Cf. Morin (1989), Blache (1990), Blache & Morin (1990) sur différentes notions de *famille de règles*.

IDENTITÉ :	$\left[\begin{array}{l} \text{NOM : } n \\ \text{INDEX : } i \\ \text{FAMILLE : } f \end{array} \right]$
FONCTION :	$\left\{ \begin{array}{l} \text{décomposition} \\ \text{adjonction} \\ \text{coordination} \end{array} \right\}$
CONTENU :	$\left[\begin{array}{l} \text{MÈRE} \quad p \\ \text{FILLES :} \quad \left[\begin{array}{l} \text{FILLES-TÊTES : } t \text{ ([NIVEAU : lexical])} \\ \text{FILLES-COMPLÈMENTS : } c \\ \text{FILLES-SPÉCIFICATEURS : } s \\ \text{FILLES-ADJOINTES : } a \end{array} \right] \end{array} \right]$

Les GRAMMAIRES D'OBJETS constituent donc un cadre où différents types de connaissances grammaticales (correspondant, par exemple, à différentes théories linguistiques) peuvent être exprimés. L'élaboration de tels cadres de *représentation par objets* de la connaissance grammaticale (les GRAMMAIRES D'OBJETS) constituera un des objectifs fondamentaux de la recherche en génie linguistique dans les années à venir.

Contraintes

Représentations à base de contraintes : grammaires de règles et grammaires de principes

On oppose souvent les théories grammaticales *basées sur des règles* (comme GSG ou GLF) et les théories *basées sur des principes* (comme PP)²¹. En fait, les grammaires de principes, tout comme les grammaires de règles (ainsi que les grammaires traditionnelles d'ailleurs), constituent des *représentations par contraintes* des connaissances grammaticales. L'avantage de telles représentations est qu'elles permettent l'abstraction sur les objets et leur représentation. On peut exprimer des contraintes sans avoir à faire référence aux objets spécifiques qui leur seront soumis. Les contraintes s'appliquent à des variables qui n'ont pas à être instanciées, mais uniquement typées abstraitement²².

Cependant, les représentations par contraintes posent de délicats problèmes de cohérence et de complexité. Comment comparer, par exemple, les hypothèses sur le

²¹ En fait, cette opposition est plus une opposition méthodologique qu'une opposition conceptuelle. *Méthodologie ascendante* : du spécifique au général dans le cas des *grammaires de règles* vs *méthodologie descendante* : du général au spécifique, dans le cas des *grammaires de principes*. La notion de *principe* est malheureusement assez floue et sert parfois à masquer une absence de rigueur et une confusion entre spécifique (qui s'oppose à général) et explicite (qui s'oppose à implicite). La concision, la clarté ou la simplicité peuvent constituer des qualités d'une description, qui peut entraîner un peu d'implicite. Mais l'implicite n'est jamais une qualité en soi (du moins en science).

²² D'ailleurs, toute contrainte constitue elle-même une fonction de type abstrait.

liage²³ d'une théorie comme PP, et celles d'une théorie comme la GSG²⁴ ? Il faudrait pouvoir définir les théories en termes de contraintes sur certains *domaines* définissables indépendamment de la théorie. C'est ce qu'ont fait Miller (1990) pour les *systèmes de liage*, Gazdar et al. (1988) pour les *théories catégorielles*, Pullum et Barker (1990) pour les relations de *commande* et nous-même pour différents domaines syntaxiques (Morin 1988, 1989, 1996).

En traitement des langues naturelles, cette nécessité théorique s'associe à des exigences pratiques. Idéalement, on voudrait pouvoir pratiquer un certain éclectisme éclairé, qui permettrait d'importer dans un système **S** les contraintes exprimées par la théorie **A** sur un domaine **D**, de les combiner avec les contraintes exprimées par la théorie **B** sur un domaine **D'** (non nécessairement entièrement distinct de **D**) et avec les contraintes exprimées par la théorie **C** sur le domaine **D''** et ainsi de suite.

Au niveau syntaxique, on pourrait, par exemple, disposer

- (a) d'une représentation relativement adéquate des phénomènes **DV** de DIATHÈSE VERBALE en français (les «voix» de la grammaire traditionnelle : ACTIVE, PASSIVE, PRONOMINALE, NEUTRE, MOYENNE, CAUSATIVE) exprimée dans le cadre de la *grammaire lexicale-fonctionnelle GLF*,
 - (b) d'une représentation des phénomènes de CLITICISATION PRONOMINALE **CP** dans le cadre de la théorie *gouvernement-liage PP*
- et
- (c) d'une représentation des phénomènes de COOCCURRENCE et d'ORDONNANCEMENT DES COMPLÉMENTS **COC** dans le cadre de la théorie de la *grammaire syntagmatique généralisée GSG*.

²³ Ou plus précisément, sur le *liage-filtrage*. Sur cette notion de *liage-filtrage* (ensemble des liens globaux entre objets présents et objets «absents» —filtrés— dans la structure), cf. Morin (1989).

²⁴ En PP, les phénomènes sont décrits au moyen de l'interaction d'un réseau de notions extrêmement complexe et mouvant d'un auteur à l'autre : *chaînes, gouvernement, K-marquage, θ -marquage, filtres, principe de projection, θ -critère, A et \neg A-positions*, etc. En GSG, on fait appel à des principes universels de propagation de valeurs (essentiellement le *principe de traits de pied* (Gazdar et al. 1985) ou le *principe de traits de liage* (Sag & Pollard 1987)) associés à un système de traits typés et à des restrictions intracatégorielles (restrictions de cooccurrence de traits, spécifications de traits par défaut) et extracatégorielles (DI-règles) sur les valeurs de certains traits. La comparaison de la théorie du liage PP et de celle de GSE est beaucoup plus facile, puisque la théorie du liage de GSE (Pollard & Sag 1994, ch. 6) est explicitement construite à partir de celle de PP, mais en utilisant la structure informationnelle et la hiérarchie d'oblicité (o-commande) plutôt que des notions configurationnelles (c-commande).

Dès lors, on pourrait vouloir intégrer ces trois fragments dans un système S :

(10)

$$S = s(\text{GLF}(\text{DV})) \quad \& \quad s(\text{PP}(\text{CP})) \quad \& \quad s(\text{GSG}(\text{COC}))^{25}$$

CONCLUSION

Nous avons examiné ici quelques-uns des problèmes liés à l'intégration des théories et des descriptions linguistiques en traitement des langues naturelles. Nous avons montré les avantages d'une approche par objets et contraintes et jeté les bases de ce que pourraient être des *institutions grammaticales*. La définition plus précise de ces domaines que constituent les institutions grammaticales, l'interprétation de fragments de descriptions dans ces termes et l'élaboration de grammaires d'objets et de contraintes constitueront les objectifs fondamentaux de notre recherche en ce qui concerne l'exploitation des théories et descriptions linguistiques en TAO.

RÉFÉRENCES

- ABEILLÉ, Anne (1993) : *Les nouvelles syntaxes*, Paris, Armand Colin.
- ABEILLÉ, Anne, GODARD, Danièle et Philip MILLER (à paraître) : *The Major Syntactic Structures of French*, Notes de cours d'ESSLLI-97, Aix-en-Provence.
- ARNOLD, D. et L. des TOMBE (1987) : «Basic theory and methodology in EUROTRA», in Nirenburg, S. (1987 réd.) *Machine Translation, Theoretical and Methodological Issues*, Cambridge, Cambridge University Press, pp. 114-135.
- BLACHE, Ph. (1990) : *L'analyse syntaxique dans le cadre des grammaires syntagmatiques généralisées, interprétations et stratégies*, Thèse de doctorat inédite, Faculté des Sciences de Luminy, Université d'Aix-Marseille II.
- BLACHE, Ph. et J.-Y. MORIN (1990) : «Bottom-Up Filtering : A Parsing Strategy for ID/LP Grammars», in Karlsson, F. (1990 réd.) *COLING-90, Proceedings*, Helsinki, Université d'Helsinki et New-York, Association for Computational Linguistics.
- BOUCHARD, L., EMIRKIANIAN, L. et J.-Y. MORIN (1991) : «Détermination de la couverture dans les interfaces en langue naturelle», in Gauthier, G. (1991 réd.) *ICO-91*.
- CHOMSKY, N. (1981) : *Lectures on Government and Binding*, Dordrecht, Foris.
- CHOMSKY, N. (1995) : *The Minimalist Program*, Cambridge, Mass., MIT Press.
- DA SYLVA, L. (1990) : *Un parseur inspiré de la théorie gouvernement-liage*, Mémoire de maîtrise inédit, Département de linguistique et de traduction, Université de Montréal.

²⁵ s est ici une fonction générique d'interprétation de l'ensemble $\mathbf{t}(\mathbf{d})$ des énoncés de la théorie \mathbf{t} sur le domaine \mathbf{d} .

- DAMOURETTE, Jacques et Édouard PICHON (1911-1940) : *Des mots à la pensée. Essai de grammaire de la langue française*, 7 volumes et un volume d'annexes, Paris, d'Artrey.
- EPSTEIN, S. et al. (1996 éd.) : *Minimal Ideas*, Amsterdam, John Benjamins.
- FODOR, J. & S. CRAIN (1990) : «Phrase Structure Parameters», *Linguistics and Philosophy*, 13, 6, pp. 619-659.
- GAZDAR, G., KLEIN, E. PULLUM, G. et I. SAG (1985) : *Generalized Phrase Structure Grammar*, Oxford, Blackwell et Cambridge, MA, Harvard Univ. Press.
- GAZDAR, G., PULLUM, G. CARPENTER, R. KLEIN, E., HUKARI, T. & R. LEVINE (1988) : «Category structures», *Computational Linguistics*, 14, 1, 1-19. (Traduction française dans Torris, T. et Ph. Miller (1990 éd.) *Formalismes syntaxiques pour le traitement automatique du langage naturel*, Paris, Hermès).
- GOGUEN, J.A. et R.M. BURSTALL (1983) : «Introducing Institutions», in Clarke, E. et D. Kozen (1983 éd.) *Logics of Programs*, Lecture Notes in Computer Science, 164, Berlin, Springer-Verlag, pp. 221-256.
- GOGUEN, J.A. et R.M. BURSTALL (1985) : *Institutions : Abstract Model Theory for Computer Science*, Rapport CSLI-86-54, Stanford, Center for the Study of Language and Information, Stanford University.
- GOGUEN, J.A. et R.M. BURSTALL (1986) : *A Study in the Foundations of Programming Methodology : Specifications, Institutions, Charters and Parchments*, Rapport CSLI-85-30, Stanford, Center for the Study of Language and Information, Stanford University.
- GROSS, Maurice (1975) : *Méthodes en syntaxe*, Paris, Hermann.
- JACKENDOFF, R. S. (1987) : *Consciousness and the Computational Mind*, Cambridge, MA, MIT Press.
- MEL'CUK, I. et al. (1984, 1988, 1992) : *Dictionnaire explicatif et combinatoire du français, contemporain Recherches lexico-sémantiques I-III*, Montréal, Presses de l'Université de Montréal.
- MILIČEVIĆ, J. (1997) : *Étiquettes sémantiques dans un dictionnaire formalisé du type Dictionnaire Explicatif et Combinatoire*, Mémoire de maîtrise, Département de linguistique et de traduction, Université de Montréal.
- MILLER, Ph. (1990) : «Systèmes de liage», in Torris, T. et Ph. Miller (1990 éd.) *Formalismes syntaxiques pour le traitement automatique du langage naturel*, Paris, Hermès.
- MILNER, J.-C. (1989) : *Introduction à une science du langage*, Paris, Seuil.
- MORIN, J.-Y. (1988) : «Prédicats théoriques et données externes. Syntaxe diachronique», *Revue canadienne de linguistique*, numéro spécial, *Linguistic Theory and External Evidence*, 33, 4, pp. 443-475.
- MORIN, J.-Y. (1989) : *Syntaxe*, Département de linguistique et philologie, Université de Montréal.
- MORIN, J.-Y. (1996) : «Configuration vs. information. An informational explanation of command relations», in Park, B.-S. & J.-B. Kim (1996 éd.) *PACLIC-11, Language, Information and Computation*, Séoul, LERI, Kyung Hee University, pp. 11-20.

- POLLARD, C.J. (1984) : *Generalized Phrase Structure Grammars, Head Grammars and Natural Language*, Thèse de Ph. D. inédite, Stanford University.
- POLLARD, C.J. & I. SAG (1994) : *Head-Driven Phrase Structure Grammar*, Chicago, Univ. of Chicago Press.
- PULLUM, G. et C. BARKER (1990) : «A Theory of Command Relations», *Natural Language and Linguistic Theory*.
- SAG, I. et C.J. POLLARD (1987) : *Information-Based Syntax and Semantics, Vol. I : Fundamentals*, Stanford, CSLI et Chicago University of Chicago Press.
- SAINT-DIZIER, P. (1989) : «Programming in logic with constraints for natural language processing», *ACL Europe 4*, pp. 87-94.
- SAINT-DIZIER, P. & P. SZPAKOWICZ (1989 réd.) : *Logic Programming and Logic Grammars*, Londres, Ellis Horwood.
- TRAVIS, L. (1989) : «Parameters of Phrase Structure», in Baltin, M. & A. Kroch (1989 réd.) *Alternative Conceptions of Phrase Structure*, Chicago, University of Chicago Press.
- ZWICKY, A. (1986) : *Interfaces*, Columbus, Ohio State University Working Papers in Linguistics, # 32.
- ZWICKY, A. (1988) : «Direct reference to Heads», *Folia Linguistica*, 22.3-4, pp. 397-404.

LA MÉMOIRE DES PARTICIPES PRÉSENT ET PASSÉ

Poul Søren KJÆRSGAARD

Universite d'Odense, Odense, Danmark

1. INTRODUCTION

La présente contribution a pour but de faire le point sur un projet de recherche¹ qui examine les participes présent et passé en danois.

D'un point de vue immanent, il s'agit de décrire leur structure (simple ou composée), leur(s) fonction(s) ainsi que des critères permettant de les distinguer des adjectifs. Il s'agira aussi d'expliquer les contraintes régissant leur usage.

D'un point de vue contrastif, le projet s'est assigné le but de décrire la distribution des participes dans des langues voisines (les langues germaniques), mais aussi de décrire leurs équivalents dans d'autres familles de langues, notamment les langues romanes. Cette description déboucherait, à terme, sur l'établissement de procédés et de règles de traduction entre ces langues.

Dans cet article, nous allons prendre, dans un premier temps, comme point de départ les participes présent et passé en français pour mettre en lumière des comportements qui s'expliquent, nous semble-t-il, en faisant référence à l'idée d'une mémoire des mots. Cette expression est bien entendu à comprendre au sens figuré. Par mémoire, nous entendons le fait que des comportements de certains mots, par exemple leur combinaison avec d'autres mots, trouvent une explication si l'on admet l'existence de traces ou de mémoires du comportement dans un état antérieur. Concrètement, certains comportements des participes présent et passé nous paraissent explicables si l'on suppose l'existence d'un lien entre ces formes et des formes verbales finies ou infinitives. Nous présumons donc que les participes sont dérivés de ces autres mots.

Le second volet de l'article examinera les participes présent et passé en danois, cette langue choisie comme représentante des langues germaniques. Là, nous constatons également qu'une description de leur structure et de leur fonction conduit à l'idée d'une

¹ Le projet décrit ici constitue avec cinq autres projets un projet d'ensemble, *Udforskning af dansk ordforråd og grammatik*, UDOG [Recherches sur le vocabulaire et la grammaire danois] visant à décrire le vocabulaire et la grammaire (syntaxe) de la langue danoise contemporaine. Il est coordonné par *Center for sprogteknologi* (Centre for Language Technology) de Copenhague et a été subventionné par le CNRS danois sous contrat n° 15-9018 ainsi que par les universités des six équipes.

mémoire de ces mots. Puisqu'il s'agit d'une autre langue, ce sont évidemment d'autres mécanismes qui entrent en jeu. Nous allons voir que la structure des participes en danois et en français conduit à des problèmes de traduction. Ces problèmes, qu'il s'agisse d'un traducteur humain ou non, deviennent explicites, si l'on s'appuie sur l'idée de mémoire, c.-à-d. de relations et de traits morphosyntaxiques, qui, certes, sont neutralisés en surface, mais qui déterminent, partiellement au moins, la combinatoire des participes.

Enfin dans la troisième partie de l'article, nous allons passer en revue l'éventail des structures utilisées en français pour rendre les constructions danoises.

2. LES PARTICIPES FRANÇAIS

2.1 Le participe présent

Le participe présent s'emploie sous deux formes différentes, l'une appelée participe présent, l'autre adjectif (verbal)². Sans entrer dans le détail — nous renvoyons aux descriptions des manuels de grammaire, p.ex. Arrivé et al. (1986 : 472-473), Riegel et al. (1997 : 340-341) et de l'article de Halmøy (1984, 52-57ss.), il existe des critères morphologiques, syntaxiques, et lexico-sémantiques, qui permettent de distinguer l'un et l'autre. Résumons-les brièvement :

- a) Le participe maintient l'orthographe [qu/gu]-ant, alors que certains adjectifs (verbaux) choisissent l'orthographe [c/g]-ant. Le participe conserve aussi *-ant*, alors que certains adjectifs (verbaux) prennent *-ent*. En plus, le participe reste inchangé, alors que l'adjectif (verbal) s'accorde en genre et en nombre avec le nom auquel il se rapporte.
- b) Le participe se combine avec les compléments en fonction de la valence et de la construction du verbe correspondant. Cette option n'existe pas pour l'adjectif (verbal). Le participe ne s'emploie pas comme attribut (du sujet), ce qui est le cas de l'adjectif (verbal)³. Les deux s'emploient comme épithète. Le participe ne se renforce pas par *très*, ce qui est le cas de l'adjectif (verbal) : *Un projet de loi (*très) intéressant les syndicats* vs *Un projet de loi (très) intéressant* (cité d'après Kjærsgaard, 1995 : 118). Épithète, le participe suivi de ses compléments est placé derrière le nom. Dans cette fonction et dans un style plutôt soutenu, l'adjectif (verbal) peut être antéposé au nom. Cette différence s'accompagne souvent d'une différence de sens, cf. ci-dessous.

² Dans un article intitulé *À propos de l'adjectif en -ant, dit "verbal"*, Odile Halmøy s'insurge contre cette étiquette. Elle propose, avec raison, de considérer la différence des deux formes comme une instance de la différence entre flexion et dérivation. Afin de marquer ce parti pris, nous avons adopté l'étiquette *adjectif (verbal)*.

³ Halmøy (1994 : 53) observe que certains adjectifs (verbaux) entrent en collocation avec certains noms, de façon à ce que la fonction attributive entre eux cesse d'exister : *eau courante* *→ *cette est courante*.

- c) Le participe des verbes perfectifs⁴ dénote l'aspect inaccompli de l'action du verbe et la simultanéité avec le verbe fini de la phrase, alors que l'adjectif (verbal) décrit plutôt un état ou une propriété du nom auquel il se rapporte.

Le participe est formé du suffixe *-ant* accolé à la racine de l'infinitif, ce qui n'est pas toujours le cas de l'adjectif (verbal). C'est ainsi qu'on a en français le participe *connaissant*, p.ex. dans *des étudiants connaissant bien la langue française*. L'adjectif (verbal) correspondant, par contre, n'est pas ce lexème, mais *connaisseur*. On n'a donc pas **des étudiants connaissants de la langue française*, mais bien *des étudiants connaisseurs de la langue française*⁵. Par contre, on a en français l'adjectif (verbal) *reconnaissant*, p.ex. dans *la patrie reconnaissante*. Ce lexème s'emploie aussi comme participe, p.ex. *les Français reconnaissant le besoin de soutenir leur langue...*

À l'inverse, et pour des raisons inexpliquées, *reconnaissseur* est inexistant.

Notons enfin des différences de sens entre le participe et l'adjectif (verbal). Dans ces cas, c'est le participe qui conserve le sens de l'infinitif, alors que l'adjectif (verbal) change de sens. Halmøy (1984 : 59) donne l'exemple de *regardant*, avec le sens de *regarder* dans *un homme regardant la télé* vs *un homme regardant* où l'adjectif (verbal) prend la valeur de *avare*. D'autres exemples cités dans les manuels de grammaire sont *violant/violent*, *comptant/comptant* [*argent comptant*].

La relation syntaxique entre un nom et un participe/adjectif (verbal) est normalement celle existant entre le sujet et le verbe (fini) : *des projets intéressant l'ensemble des professeurs* vs *des projets intéressants*. Cette relation apparaît en paraphrasant l'expression à l'aide d'une proposition relative : *des projets qui intéressent l'ensemble des professeurs* vs *des projets qui intéressent/sont intéressants*⁶. La relation sémantique entre un nom et un participe/adjectif (verbal) est active ou passive selon le sens du verbe : *les chiens aboyants* vs *les nations naissantes*.

Épithète, l'adjectif (verbal) autorise parfois un relâchement des contraintes valencielles : *les villes environnantes*, alors qu'on ne trouverait guère le verbe *environner* sans complément d'objet direct.

C'est dans ce contexte qu'on trouve une première instance de mémoire, car l'adjectif (verbal) entretient parfois des relations diverses avec la tête du groupe nominal dans lequel il se trouve.

Exemples cités d'après Arrivé et al. (1984 : 473) et Riegel et al. (1997 : 341) :

<i>une couleur voyante</i>	→ <i>une couleur qu'on voit (très bien)</i>	sens passif
<i>une personne méfiante</i>	→ <i>une personne qui se méfie</i>	sens réfléchi

⁴ Cette distinction ne s'applique pas aux verbes imperfectifs (cf. Halmøy, 1984 : 55). Un exemple en est le verbe *surveiller* : *des femmes surveillant le square* vs *des femmes surveillantes*.

⁵ D'autres exemples sont présentés par Halmøy (1984 : 49) : *un site *enchantant/enchanteur*, *un calme *trompant/trompeur*, *un siège *inclinant/inclinable*. Dans ces cas, le participe serait p.ex. *un siège s'inclinant en arrière*, *un calme trompant l'entourage*.

⁶ Cette paraphrase ne constitue pas une relation bijective, c.-à-d. elle ne s'applique pas toujours dans les deux sens comme nous l'avons vu ci-dessus : *un site qui enchante* → *un site *enchantant*.

<i>un revêtement glissant</i>	→ <i>un revêtement qui fait glisser</i>	sens factitif
<i>une soirée dansante</i>	→ <i>une soirée durant laquelle on danse</i>	sens temporel
<i>une rue passante</i>	→ <i>une rue où beaucoup de gens passent</i>	sens locatif.

Ces exemples nous font dire que les adjectifs (verbaux) sont susceptibles de recouvrir des valeurs neutralisées, mais non pas disparues, car elles demeurent à l'état virtuel et réapparaissent au paraphrasage.

2.2 Le participe passé

Le participe passé s'emploie, comme le participe présent, de deux façons différentes, soit comme tête d'un groupe adjectival sans compléments (équivalent de l'adjectif (verbal)), soit comme tête d'un groupe adjectival avec compléments (équivalent du participe). La principale différence entre elles est, comme pour le participe présent, la gamme réduite de fonctions du dernier qui ne s'emploie guère comme attribut (du sujet) : *des chambres réservées* vs *des chambres sont réservées*, mais *des chambres à eux⁷ réservées* vs *des chambres sont *[à eux] réservées*.

Contrairement à son homologue, le participe passé connaît cependant des contraintes de combinatoire et de formation.

Épithète d'un groupe nominal, le participe n'est en principe compatible qu'avec une tête qui fonctionne comme le complément d'objet direct du verbe fini. Cette règle constitue la contrepartie logique de la règle employée pour le participe présent, cf. ci-dessus.

Cela explique qu'on a *les paquets expédiés* (quelqu'un a expédié les paquets), mais non pas **les étudiants expédiés* (quelqu'un a expédié les paquets aux étudiants).

La règle connaît au moins deux entorses : quelle que soit la fonction syntaxique du nom, le rôle sémantique de celui-ci doit être celui d'un patient ou d'un thème (selon la typologie et la taxinomie employées). Cela explique l'existence de *les enfants morts pour la patrie*. Dans cet exemple, la relation syntaxique entre le nom et le participe épithète est celle d'un sujet et d'un verbe, mais leur relation sémantique est celle qui existe entre un verbe et un patient. La relation sémantique prime donc la relation syntaxique.

Comme pour le participe présent, il existe aussi des collocations dans lesquelles les rapports normaux sont suspendus ou dans lesquels la thèse d'une relation fixe (verbe-patient) n'a pas de sens. Mentionnons (d'après Riegel, 1997 : 344) *le journal parlé, la presse écrite, l'eau salée*.

Pour ce dernier exemple, on remarque qu'il est passé complètement dans la catégorie des adjectifs, un critère valable étant sa compatibilité avec *très*⁸.

⁷ *à eux* s'analyse comme un complément d'attribution (objet datif).

⁸ L'auteur de cet article a esquissé une série de tests permettant de distinguer adjectifs et participes déverbaux en danois : Kjærsgaard et le Fevre Jakobsen (1997, à paraître).

Faisant abstraction des contre-exemples, il semble légitime d'affirmer que la combinaison d'un nom et d'un participe passé épithète dans la grande majorité des cas est régie (et contrainte) par des relations qui se trouvent sur un niveau différent de l'analyse, celle de la phrase ou celle des relations verbe-compléments. De là à affirmer que le participe passé épithète se souvient de (ou a mémorisé des) comportements inhérents au verbe, il n'y a qu'un pas.

Le participe passé connaît aussi une autre contrainte importante à son emploi comme épithète. Alors que l'adjectif (verbal) s'emploie comme épithète des verbes monovalents (intransitifs) inergatifs⁹, il n'en est rien du participe passé :

<i>eau courante</i>	<i>*eau courue</i>
<i>eau dormante</i>	<i>*eau dormie</i>
<i>chien aboyant</i>	<i>*chien aboyé</i>

L'emploi épithétique existe pour les deux participes des verbes monovalents ergatifs :

<i>les enfants mourants</i>	<i>les enfants morts</i>
<i>le caramel fondant</i>	<i>le caramel fondu</i>

La conclusion qui s'impose est que le comportement syntaxique des participes épithètes, y compris leurs /in-/compatibilités avec la tête du groupe nominal, reflète un embryon de mémoire des propriétés valables pour les verbes finis. On peut exprimer la même conclusion en disant que les contraintes valenciennes (nombre des actants, nature de leurs constituants, leurs fonctions syntaxiques, leurs rôles sémantiques et les contraintes de sélections qui y sont associées) qui régissent les rapports entre le verbe fini et ses actants (sujet + compléments) sont mémorisés par les participes. Enfin, c'est un savoir intrinsèque qui demeure à l'état virtuel, récupérable au besoin.

3. LES PARTICIPES DANOIS

3.1 Les participes simples

Les participes danois sont employés moins fréquemment que leurs homologues français. Cela tient à différentes raisons que la présente contribution n'examinera pas.

Un participe présent danois se traduit parfois par un participe passé français ou vice versa :

<i>et foretagende</i> (art. indéf.neut.sg.-part.pr. substantivé)	<i>une entreprise</i>
<i>en tilbundsgående undersøgelse</i>	<i>une enquête approfondie</i>

⁹ Les expressions *inergatif* et *ergatif* s'emploient selon la terminologie de la grammaire relationnelle, développée par David Perlmutter et autres. Un verbe inergatif est un verbe (intransitif) à la voix active dont l'action part du sujet : *courir*. Un verbe ergatif (ou inaccusatif) est un verbe (intransitif) à la voix active dont l'action atteint ou vise le sujet : *périr*.

(art. indéf.com.sg.¹⁰-part.pr.-nom sg.)

en befærdet gade *une rue passante*
 (art. indéf.com.sg.-part.pa.-nom sg.)

Ces différences lexicales sont considérées comme des cas de l'arbitraire du signifiant et ne seront pas non plus traitées dans le cadre présent.

Les relations sémantico-syntaxiques qui valent pour les participes français, sont valables grosso modo pour les participes danois.

Le rapport normal entre un nom et un participe présent/adjectif (verbal) épithète est celui du verbe et de son sujet. Parallèlement, la relation qui existe entre un nom et un participe passé épithète correspond à celle qu'on trouve entre le verbe et son objet direct. On voit donc en danois, grosso modo, les mêmes contraintes qu'en français.

Pour les monovalents inergatifs, le danois se comporte comme le français¹¹ : le participe présent existe, alors que ce n'est pas le cas du participe passé :

de løbende atleter *les athlètes qui courent*¹²
 (art. déf.pl.-part.pr.:courir-nom pl.:athlète)

*de *løbne atleter* **les athlètes courus*
 (art. déf.pl.-part.pa.:courir-nom pl.:athlète)

de knoklende landmænd *les paysans qui bossent*
 (art. déf.-part.pr.:bossier-nom pl.:paysan)

*de *knoklede landmænd* **les paysans bossés*

Les monovalents ergatifs (inaccusatifs) connaissent les deux participes :

det synkende skib *le bateau qui fait naufrage*
 (art. def.sg.neut.-part.pr.:faire naufrage-nom sg.:bateau)

det sunkne skib *le bateau qui a fait naufrage*
 (art. def.sg.-part.pa.:faire naufrage-nom sg.:bateau)

de sammenstyrtende huse *les maisons qui s'écroulent*
 (art. def.pl.-part.pr.:s'écrouler-nom pl.:maison)

¹⁰ Le danois possède deux genres : le neutre et le commun (amalgame du masculin et du féminin).

¹¹ Les langues française et danoise ne sont pas les seules à se comporter ainsi. À ma connaissance, la grammaire relationnelle pose cette relation comme universelle.

¹² Les traductions fournies sont celles qui paraissent le plus appropriées pour une traduction hors contexte. C'est ainsi qu'un participe présent danois équivaut souvent à une relative française. Le contenu des parenthèses indiquent la partie du discours des mots danois.

de sammenstyrtede huse *les maisons écroulées/qui se sont écroulées*
(art. déf.pl.-part.pa.:s'écrouler-nom pl.:maison)

Dans ces cas, la différence demeure partout celle qui existe entre une acception aspectuelle inaccomplie (simultanée) et une autre accomplie.

Pour les verbes bivalents inergatifs, on observe les patterns suivants :

*de spisende børn*¹³ *les enfants qui mangent*
(art. déf.pl.-part.pr.:manger-nom pl.:enfant)

de spiste grønsager *les haricots consommés*
(art. déf.pl.-part.pa.:consommer-nom pl.:haricot)

de skabende erhverv *les métiers créateurs*¹⁴
(art. déf.pl.-part.pr.:créer-nom pl.:métier)

de skabte værdier *les plus-values créées*
(art. déf.pl.-part.pa.:créer-nom pl.:valeur)

Dans aucun exemple cité, il ne serait possible de substituer un participe présent à un participe passé ou vice versa. Cela s'explique par des contraintes valencielles au niveau sémantique des verbes en question. Les verbes *spise* (manger, consommer) et *skabe* (créer) supposent normalement deux actants, dont le premier est sujet et animé et le second est objet direct et inanimé.

Le même pattern s'applique aux verbes ergatifs :

de lidende børn *les enfants qui souffrent*
(art. déf.pl.-part.pr.:souffrir-nom pl.:enfant)

de lidte tab *les pertes subies*
(art. déf.pl.-part.pa.:souffrir-nom pl.:perte)

¹³ Dans les exemples suivants, un actant, facultatif, est omis. Pour exprimer un nom d'agent, le danois peut même omettre tous les actants. C'est ainsi qu'on a *de spisende* (participe présent substantivé au plur.déf.) qui correspond à *les mangeurs*. Il existe une concurrence entre le nom d'agent et le participe présent substantivé, apparemment en danois comme en français : *de levende* vs *les vivants* (tous deux des participes); *vinderne* (plur.déf.) vs *les gagnants* (nom d'agent vs participe présent). Aucune règle ne permet de prédire si c'est l'une ou l'autre construction qui l'emporte (l'arbitraire du signifiant).

Pour les verbes trivalents, assez rares, le nom d'agent l'emporte sur le participe présent substantivé : *giverne* (nom-art.déf.pl. accolé au nom) - *les donneurs*; *lånerne* - *les emprunteurs/les prêteurs*.

¹⁴ La différence entre un participe présent tête d'un groupe à compléments et un adjectif (verbal) ne se manifeste pas au niveau lexical comme c'est parfois le cas en français, cf. par. 2.1.

3.2 Les participes composés

Les langues germaniques, dont le danois, se caractérisent encore par un phénomène quasiment inconnu des langues romanes : la faculté de préfixation d'un lexème (nom, adjectif, adverbe ou particule), le nouvel ensemble constituant un composé synthétique et assumant les mêmes fonctions syntaxiques que le participe simple, c.-à-d. principalement comme épithète d'un nom tête d'un groupe nominal ou comme attribut¹⁵. Il s'agit d'un processus de formation lexicale très productif en danois contemporain.

Cette construction synthétique existe aussi pour les noms et les autres parties du discours. Pour des raisons qui ne seront expliquées ici, il ne peut être préfixé qu'un seul lexème au participe (ou, comme nous l'appellerons dorénavant : participe déverbal (pour les distinguer des adjectifs)). Cela signifie que, pour les verbes trivalents, au moins un actant ne peut être exprimé.

Pour les participes présent et passé, le lexème préfixé et le nom tête du groupe nominal auquel se rapporte le participe déverbal, assument des rapports qui se résument par deux patterns archétypiques :

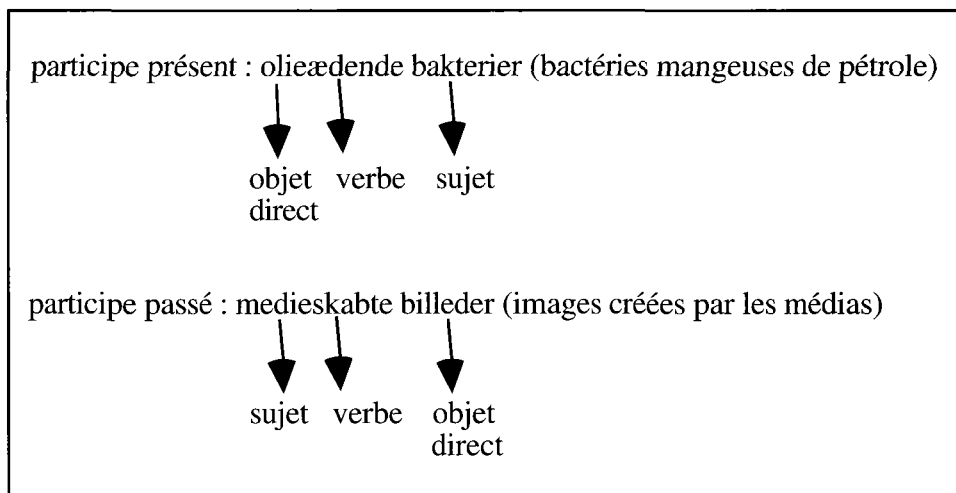


Figure 1

¹⁵ Les exceptions des langues romanes sont peu nombreuses et se répartissent en deux groupes :

a) Mots d'emprunt d'origine grecque ou latine qui correspondent aux participes danois (et a fortiori germaniques) :

cancérogène vs *kræftfremkaldende* [nom:cancer-part.pr.:provoquer];

bactéricide vs *bakterieædende* [nom:bactérie-part.pr.:manger].

b) Hapax, que je sache, en espagnol : *hispanohablante* vs *spansktalende*.

Ajoutons enfin que tel participe composé danois (ou germanique) correspond parfois à un lexème non composé en français : *computerstyret* [nom:ordinateur-part.pa.:diriger/exécuter] vs *informatisé*.

Ce sont ces patterns qui permettent de prédire la possibilité des constructions suivantes :

laboratoriekontrollerede fødevarer des aliments contrôlés par des laboratoires
([nom:laboratoire-part. pa.:contrôler]-nom pl.:aliment)

fødevarekontrollerende laboratorier des laboratoires contrôlant les aliments
([nom:aliment-part.pr.:contrôler]-nom pl.:laboratoire)

Ce sont aussi ces patterns qui permettent de prédire l'impossibilité des constructions inverses :

**fødevarekontrollerede laboratorier* *des laboratoires contrôlés par des aliments
(nom:aliment-part.pa.:contrôler]-nom pl.:laboratoire)

**laboratoriekontrollerende fødevarer* *des aliments contrôlant les laboratoires
([nom:laboratoire-part.pr.:contrôler]-nom pl.:aliment)

Les contraintes de sélection sémantiques du verbe *kontrollere* [contrôler] interdisent tout simplement ces constructions. Voici un nouvel élément qui, nous semble-t-il, justifie l'affirmation que le participe déverbal, dérivé du verbe, conserve un embryon de mémoire.

Aux fonctions syntaxiques indiquées s'ajoutent évidemment des rôles sémantiques : le sujet correspondant le plus souvent à l'agent et l'objet direct au patient/thème.

Le schéma reproduit ci-dessus est pourtant loin de représenter l'intégralité de la réalité linguistique. Si le participe déverbal dérive d'un verbe régissant un complément indirect (objet prépositionnel), la préposition est supprimée pendant le processus de préfixation, et structurellement, il n'y a donc pas moyen de voir de différence entre des objets direct et indirect préfixés. Les compléments circonstanciels, souvent exprimés à l'aide d'un groupe prépositionnel, se voient aussi privés de préposition : Il s'ensuit qu'on ignore la fonction et le rôle du préfixe (cas à distinguer : sujet ou complément circonstanciel accolé au participe passé; objet direct ou complément circonstanciel accolé au participe présent). Par conséquent, les constructions préfixées sont souvent structurellement ambiguës (voir exemple ci-dessous).

Puisque la plupart des participes déverbaux ne sont pas (et ne devraient pas être) lexicalisés dans les dictionnaires (ni pour humains, ni pour ordinateurs), il s'ensuit que ces ambiguïtés ne sont pas directement décelables, et que la détermination des relations entre le participe déverbal d'un côté, et l'élément préfixé et le nom tête du groupe nominal de l'autre, requièrent une analyse préalable. L'exemple suivant illustre cette ambiguïté :

vandforurenende virksomheder des entreprises polluant l'eau
([nom: eau-part.pr.:polluer]-nom pl.:entreprise)

olieforurenende virksomheder des entreprises polluant par le pétrole
([nom: pétrole-part.pr.:polluer]-nom pl.:entreprise)

La traduction vers le français indique la diversité des rôles sémantiques et fonctions syntaxiques des deux exemples. Dans le premier, le préfixe obéit au pattern archétypique du participe présent : objet direct - verbe - sujet. Dans le second, le préfixe ne saurait assumer cette fonction (ce serait contraire à notre intuition sur le fonctionnement 'normal' de notre monde), mais celle d'un complément circonstanciel ou du point de vue sémantique celle du moyen¹⁶.

Il s'est avéré pour le danois qu'une analyse, mettant en œuvre la partie du discours du préfixe, la construction verbale du verbe dont le participe déverbal est dérivé ainsi que les contraintes de sélection de celui-ci, parvient à désambiguïser la plupart des constructions examinées. Cette analyse évitera du travail et de l'espace dans les dictionnaires, car elle permet de faire l'économie de la lexicalisation de ces participes composés dont le sens est compositionnel, donc prévisible.

Ce ne sont pas seulement les actants (le préfixe et le nom tête du groupe nominal) qui mémorisent des informations nécessaires à l'analyse correcte. D'autres facteurs interviennent, notamment les traits morphosyntaxiques du préfixe et certains traits inhérents au verbe/participe déverbal. Dans les constructions examinées ici, on constate d'une part que ces informations sont neutralisées, et d'autre part qu'elles sont requises en vue d'une traduction satisfaisante.

La neutralisation consiste en ce que les informations ne sont nulle part récupérables dans l'énoncé linguistique, que celui-ci soit écrit ou oral. Il faut donc supposer leur existence à l'état virtuel, soit dans la mémoire des mots, soit dans le savoir partagé des interlocuteurs.

Nous allons aborder dans un premier temps les informations mémorisées dans le verbe, puis celles du préfixe.

Tout particule qui concourt à la construction d'un verbe fini se voit neutralisé dans la construction d'un participe déverbal. Il s'agit des prépositions des verbes transitifs indirects, comme l'exemple évoqué ci-dessus : *forurene* [polluer]. Une variante de cette neutralisation apparaît pour les verbes à particule (non accompagnés de régime). Dans le participe déverbal, la distinction susceptible de modifier le sens d'un verbe préfixé d'une particule et celui d'un verbe que ce même particule suit, est neutralisée. Ainsi, en danois, on distingue *gå på* [én] - *attaquer/agresser qn. de pågå* - *avoir lieu*. Les deux forment le même participe présent déverbal : *pågående*. Ce mot, employé comme épithète, mémorise par conséquent les deux acceptions différentes. Seul l'usage des contraintes de sélection (traits sémantiques) permet de désambiguïser correctement.

Il s'agit aussi, comme en français (cf. Riegel, 1997 : 343 (*illusion évanouie*)), du pronom réfléchi, qui, lui aussi, est neutralisé pendant le processus de préfixation.

Il s'agit, enfin, du mode d'action du verbe. Bien qu'invisible dans le signifiant linguistique, il donne lieu à différentes paraphrases ou différentes traductions. C'est ainsi

¹⁶ Le point illustré par ces exemples est peut-être plus clair encore quand il s'agit des noms équivalents : *vandforurening* (pollution de l'eau) et *olieforurening* (pollution par les hydrocarbures).

qu'un participe déverbal dérivé d'un verbe imperfectif correspond à une périphrase en présent :

sygdomstruede landmænd *des paysans que la maladie menace*
([nom:maladie-épenhèse : s-part.pa.:menacer]-nom pl.:paysan)

Un participe déverbal dérivé d'un verbe perfectif correspond, en revanche, à une périphrase en parfait :

sygdomsramte landmænd *des paysans que la maladie a atteints*
([nom:maladie-épenhèse : s-part. pa.:atteindre]-nom pl.:paysan)

À moins de préférer l'équivalent participe qui existe aussi en français [*menacés/atteints (de/par)*], ces exemples montrent que le mode d'action détermine le choix du temps du verbe.

En ce qui concerne les préfixes, ce sont la détermination, le nombre, l'accord et la (forme de la) négation qui sont neutralisées ou modifiées. Là aussi, une traduction appropriée présuppose l'ajout de ces informations. Exemples :

udlandsproducerende virksomheder *des entreprises qui produisent*
([nom:étranger-épenhèse :s-part.pr.:
produire]-nom pl.:entreprise) à *l'étranger*¹⁷
Neutralisation de la détermination

arbejds giveradministrerede aftaler *des conventions gérées par les*
([nom:patron-part.pa.:gérer]-nom pl.:
accord) *patrons*
Neutralisation du nombre et de la détermination

bomberamte tyske byer *des villes allemandes que*
([nom:bombe-part.pa.:atteindre]-adj.:
allemand-nom pl.:ville) *des bombes ont atteintes*
[*des villes all. bombardées*]

bomberamte japanske byer *des villes japonaises que la bombe a*
([nom:bombe-part.pa.:atteindre]-adj.:
japonais-nom pl.:ville) *atteintes*¹⁸
[*des villes jap. bombardées*]

Mentionnons, pour des soucis d'exhaustivité, que les neutralisations accompagnant le processus de préfixation, comprennent aussi l'amalgame de différentes parties du discours (qui se réfèrent au même signifié) ainsi que la neutralisation des contraintes de sélection qui s'appliquent normalement.

¹⁷ L'exemple montre qu'il peut y avoir neutralisation de plusieurs traits à la fois.

¹⁸ C'est notre savoir de l'histoire contemporaine ainsi que notre connaissance du contexte de l'exemple qui détermineront la bonne traduction. L'essentiel, c'est que le préfixe conserve la même forme, qu'il désigne un signifié singulier ou pluriel.

4. LA TRADUCTION DES PARTICIPES DÉVERBAUX

La conclusion qui s'impose après l'analyse des participes danois composés, c'est qu'une analyse des rapports syntaxiques et sémantiques des participes danois doit nécessairement précéder toute tentative de traduction vers une langue romane comme le français. Sinon, les participes danois, structurellement ambigus et susceptibles de contenir des informations neutralisées, risquent d'être traduits de façon inadéquate.

Dans cette section, nous allons survoler les constructions envisagées afin de présenter les principaux équivalents employés en français pour rendre les participes composés danois. Précisons tout de suite qu'il s'agira d'une simple juxtaposition commentée de quelques exemples. Une véritable analyse pouvant mener à l'établissement de règles de traduction paraît prématurée, vu l'état d'avancement actuel de cette phase du projet.

Nous allons prendre cinq exemples, relevés dans la presse écrite de ces dernières années et censés représenter, sinon de façon exhaustive, du moins une gamme considérable des constructions employées. Les exemples se répartissent en trois groupes, selon la paraphrase choisie :

a) participe déverbal danois vs groupe prépositionnel en français :

<i>det <u>atomdrevne</u> hangarskib CdG</i>	<i>le (porte-avions) Charles-de-</i>
(art. def.neut.sg.-[nom: <i>atome</i> -part.pa.: <i>propulser</i>]-nom sg.: <i>porte-avions</i>)	<i>Gaulle, à <u>propulsion nucléaire</u></i>

Dans cet exemple le préfixe danois *atom* désigne le moyen de locomotion de ce porte-avions, autrement dit le rôle sémantique. En outre, l'exemple fournit une illustration de cet amalgame de mots auquel il a été fait référence à la fin de la section 3.2. Car ce ne sont pas les atomes qui propulsent ce bâtiment, c'est l'énergie libérée à leur fission qui joue. L'expression correcte et exhaustive serait donc *atomkraftdrevne hangarskib CdG*.

b) participe déverbal danois vs proposition relative en français :

<i>den eneste tilbageværende trussel</i>	<i>la seule menace qui continue à</i>
(art. déf.sg.comm.-adj: <i>seul</i> -part.pr.: <i>rester</i> -nom sg.: <i>menace</i>)	<i>exister consiste à...</i>

C'est sans doute la paraphrase la plus fréquente et celle à laquelle il a déjà été fait allusion ci-dessus. Son emploi présuppose d'un côté l'interprétation correcte des éléments du composé et de l'autre, qu'il n'y ait pas de relatives en cascades. Ce phénomène obère l'identification de l'anaphore au point qu'on préfère souvent d'autres constructions.

c) participe déverbal danois vs groupe adjectival en français :

<i>de <u>værdiskabende</u> kilder</i>	<i>les sources <u>productrices de valeurs</u></i>
(art. déf.pl.-[nom: <i>valeur</i> -part.pr.: <i>créer</i>]-nom pl.: <i>source</i>)	

de *cadmiumforgiftede jordprøver*
(art. déf.pl.-[nom:*cadmium*-part.pa.:
contaminer]-nom pl.:*échantillon*)

les échantillons de sols *contaminés*
au cadmium

de *statsgaranterede lån til SMV*
(art. déf.pl.-[nom:*État-épenthèse*:s-
part.pa.:*garantir*]-nom pl.:*prêt*)

les prêts aux PME *garantis par*
l'État

Là où le rapport entre le participe déverbal et son préfixe est neutralisé en danois, il est partiellement explicite en français par le moyen d'une préposition. Cette clarté est pourtant toute relative, car la préposition *par*, pour prendre un exemple, peut désigner soit l'agent, comme c'est le cas de l'exemple cité, soit le moyen. L'analyse qu'une paraphrase relative fournirait demeure donc potentiellement nécessaire.

Mentionnons une autre stratégie qui consiste à ne pas traduire le préfixe. Nous ne disposons pas d'exemples de l'emploi de cette stratégie pour la traduction entre le français et le danois. Il n'y aurait sans doute pas de difficultés à en repérer. Un exemple venant d'une traduction du danois vers l'anglais fournit cependant une illustration de cette stratégie. On sait que l'action du verbe *udhule* [creuser] demande un agent (humain), un patient (un tronc) et enfin un moyen (un outil). Le nombre d'actants préfixés étant limités à un seul, on a pour le participe passé le choix entre le sujet (l'agent) ou le moyen. L'un ou l'autre est sous-entendu. Dans l'exemple suivant, le danois avait préféré préfixer l'outil, alors qu'en anglais celui-ci n'était pas exprimé :

en økseudhulet lindestamme

a hollowed-out lime trunk
[un tronc de tilleul creusé]

5. (EN GUISE DE) CONCLUSION

L'exposé qu'on vient de lire a voulu décrire une série de contraintes auxquelles est sujet l'usage des participes français et danois. Pour les participes simples, il s'agit notamment de contraintes de formation et de combinatoire.

Il semble que ces contraintes s'expliquent le mieux en faisant référence à un savoir implicite ou bien à un embryon de mémoire qui conserve les propriétés issues de la construction des verbes dont les participes sont dérivés.

Il est à noter que les deux langues sont sujettes à des contraintes qui sont souvent identiques.

Puisque la formation lexicale danoise autorise une préfixation inconnue en français, la traduction entre ces deux langues et a fortiori entre les familles dont ces deux langues sont issues, doit tenir compte d'une série de problèmes qui, eux aussi, font appel à la mémoire des mots. Il s'agit de deux groupes de problèmes :

- ambiguïté structurelle, qu'il s'agisse de la mémoire de la construction syntaxique ou de la mémoire des contraintes sémantiques;

récupération du savoir implicite contenu dans le participe déverbal ou dans son préfixe.

Une traduction entre ces deux langues exige que ces problèmes soient résolus de manière satisfaisante.

6. RÉFÉRENCES

- ARRIVÉ, Michel, GADET, Françoise et Michel GALMICHE (1986) : *La grammaire d'aujourd'hui – guide alphabétique de linguistique française*, Paris, Flammarion, 720 p.
- GREVISSE, Maurice et André GOOSSE (1993) : *Le bon usage*, Paris/Louvain-la-Neuve, DeBoeck et Duculot, 1762 p.
- HALMØY, Odile (1984) : «À propos de l'adjectif en -ant, dit "verbal"», *Revue Romane*, 19.1, Copenhague, pp. 48-64.
- KJÆRSGAARD, Poul Søren (1995) : *Fransk grammatik - i hovedtræk* [Grammaire française], Odense, Odense Universitetsforlag, 310 p.
- KJÆRSGAARD, Poul Søren (1995) : «Disambiguering af enkle og sammensatte participialer i dansk» [La désambiguïsation des participes déverbaux simples et composés en danois], Poul Søren Kjærsgaard et Lene Schøslér (Eds.), *UDOG – Udforskning af dansk ordforråd og grammatik* [Recherches sur le vocabulaire et la grammaire danois], coll. Udog, n° 3, Odense, pp. 9-22.
- KJÆRSGAARD, Poul Søren (1996) : «Danske participialer og valens» [Les participes danois et la valence], Karin Van Durme (Ed.), *Adjektivernes valens* [La valence des adjectifs], coll. Odense Working Papers in Language and Communication, n° 12, Odense, pp. 49-84.
- KJÆRSGAARD, Poul Søren et Bjarne LE FEVRE JAKOBSEN (1997, à paraître) : «Sondringen mellem participiale deverbaler og adjektiver» [La distinction entre participes déverbaux et adjectifs en danois], Bente Maegaard et Bolette Sandford Pedersen (Eds.), *UDOG – Udforskning af dansk ordforråd og grammatik* [Recherches sur le vocabulaire et la grammaire danois], coll. Udog, n° 6, Copenhague.
- PERLMUTTER, David M. and Carol G. ROSEN (Eds) (1983-90) : *Studies in Relational Grammar*, I-III, The University of Chicago, Chicago, Ill.
- PÛMPPEL-MADER, Maria, GASSNER-KOCH, Elsbeth und Hans WELLMANN (1992) : *Deutsche Wortbildung. Typen und Tendenzen in der Gegenwartssprache*, Tome 5, Adjektivkomposita und Partizipialbildungen, (Komposita und kompositionsähnliche Strukturen). Berlin/New York, Walter de Gruyter, 340 p.
- RIEGEL, Michel, PELLET, Jean-Christophe et René RIOUL (1997) : *Grammaire méthodique du français*, Paris, Presses Universitaires de France, 646 p.

Remerciements

L'auteur tient à remercier ses collaborateurs de ce projet pour leur aide et leurs commentaires durant l'élaboration de cet article, Bjarne le Fevre Jakobsen, Jens Ahlmann Hansen et Peter Stewart.

Les exemples français fournis sont, dans la mesure du possible, vérifiés à l'aide de la base textuelle de l'INaLF, Frantext.

EXPLORATION DE CLASSIFIEURS CONNEXIONNISTES POUR L'ANALYSE DE TEXTES ASSISTÉE PAR ORDINATEUR

Jean-Guy MEUNIER, Ismaïl BISKRI, Georges NAULT, Moses NYONGWA

LANCI, Université du Québec à Montréal, Montréal, Canada

1. PRÉSENTATION GÉNÉRALE

Cet article présente un volet de la recherche et du développement effectués dans le cadre du projet Franco-Québécois intitulé «Les classifieurs émergentistes et le traitement de l'information». L'objectif général de ce projet est d'explorer les approches classificatoires statistico numériques pour l'analyse de l'information se présentant en langue naturelle, texte, fiche, documents etc. (leur pertinence également). Dans le présent travail, les classifieurs explorés ont été les réseaux de neurones, les analyses multidimensionnelles, les algorithmes génétiques, les champs de Markov. Concrètement, le but visé par ces recherches est de trouver une méthode efficace et économique d'appariement de tels types de données informationnelles. Dans le cas du texte, il s'agit de formation de classes de segments textuels qui se ressemblent en raison d'un critère particulier choisi dans l'expérimentation. Ces classes peuvent alors être soumises à d'autres analyseurs qui eux ouvrent la voie à des fonctions d'analyses plus complexes. Celles-ci deviennent alors des moteurs de fouille du type hypertextualisation automatique, extraction de connaissances, indexation, analyse terminologique etc. Dans le cas de fiches descriptives (genre : dossiers de patients), il s'agit aussi de formation de classes de fiches présentant des similarités, par exemple entre patients ou entre symptômes. Ces classes sont ensuite soumises à des analyses plus fines et intégrées dans des systèmes d'assistance dynamique aux diagnostics.

Un des résultats périphériques mais des plus prometteurs de cette collaboration est l'entente explicite d'un certain nombre de chercheurs pour collaborer à un protocole ou une plate-forme commune afin de permettre aux divers logiciels francophones utilisés dans ce projet d'entrer en interaction. Trop de logiciels français restent sur les tablettes faute de plate-forme pour les mettre en interaction. Le LANCI propose une plate-forme, ALADIN (Seffah et Meunier, 1995), qui constitue une première tentative, prototypale, pour mettre concrètement en interaction ces logiciels. Les laboratoires LANCI et TIMC-IMAG travaillent dans le cadre du présent projet à la réalisation de cet objectif. Dans cet article, nous présentons brièvement l'exploration et l'expérimentation d'une approche neuronale pour l'analyse terminologique dans les grand corpus.

2. COLLABORATION FRANCO-QUÉBÉCOISE

Ce projet a été réalisé dans le cadre d'un programme Franco-Québécois, sous la co-tutelle du ministère de l'Enseignement supérieur et de la Recherche pour la partie française (M.E.N.E.S.R.I.P., Délégation à l'Information Scientifique et Technique, Programme Ingénierie linguistique et de la connaissance) et du ministère de la Recherche et du ministère des Affaires Internationales, de l'Immigration et des communautés Culturelles pour la partie Québécoise.

Du côté français, ce projet s'est déployé en une collaboration portant sur l'assistance au diagnostique médical dynamique où participent Dr. Danel, du CHU de Grenoble, V. Rialle (et son équipe) du TIMC IMAG de Grenoble.

Du côté Québécois, en raison de la dimension textuelle, ce projet a permis de dynamiser grandement l'intégration de l'équipe à plusieurs projets AUPELF-UREF FRANCIL. sur les systèmes logiciels d'assistance à la terminologie (ARC A3). Ce qui s'est à son tour déployé en une collaboration intense et suivie avec l'université de Paris-Sorbonne (Laboratoire CAMS-LALIC : dir. J.P. Desclés) ainsi que l'IDIST de Lille 3.

3. CLASSIFIEURS NEURONAUX ET ASSISTANCE À LA TERMINOLOGIE

De nos jours, un nombre croissant d'institutions accumulent très rapidement des quantités de documents qui ne sont souvent classés ou catégorisés que très sommairement. Très vite, les tâches de dépistage, d'exploration et de récupération de l'information présente dans ces textes, c'est-à-dire des «connaissances», deviennent extrêmement ardues, sinon impossibles. Pour y faire face, il devient nécessaire d'explorer de nouvelles approches d'aide à la lecture et à l'analyse de textes assistées par ordinateur (LATAO).

Extraction et classification

La littérature technique relative au traitement de l'information textuelle a montré qu'il était possible d'explorer des outils d'extraction des connaissances dans des textes (*data mining*). Pour les chercheurs dans le domaine de LATAO, cette problématique n'est pas nouvelle. Dans la recherche antérieure, plusieurs techniques et méthodes ont déjà été proposées pour tenter d'organiser le contenu d'un texte en des configurations interprétables. Ces méthodes, souvent moins fines certes que les approches linguistiques et conceptuelles n'en permettent pas moins un premier parcours général et robuste du texte. Elles sont en mesure, par exemple, d'identifier dans un corpus des classes ou des groupes de lexèmes qui entretiennent entre eux des associations dites de cooccurrence et donc de détecter leurs réseaux sémantiques. Et les recherches actuelles commencent d'ailleurs à les privilégier de plus en plus. Parmi les modèles les plus couramment utilisés, on trouve habituellement l'analyse des cooccurrences, l'analyse corrélationnelle, l'analyse en composante principale, l'analyse en groupe, l'analyse factorielle, l'analyse discriminante, etc. Malgré le succès qu'elles ont obtenu, on a dû constater que ces méthodes particulières posent deux problèmes importants.

Premièrement, les modèles classiques ne peuvent traiter que des corpus stables. Toute modification du corpus exige une reprise de l'analyse numérique. Ceci devient un problème majeur dans des situations où le corpus est en constante modification (par exemple, les reposoirs de l'autoroute électronique). Deuxièmement, les types de résultats qu'ils produisent ne sont pas sans problèmes théoriques. Ils posent des problèmes d'interprétation linguistique importants (Church et Hanks, 1990). Les associations de mots dans les classes ne sont pas toujours facilement interprétables. Pourtant, malgré leurs limites, ces approches ont été reconnues des plus utiles pour l'extraction des connaissances et plus particulièrement les connaissances terminologiques. D'une part, ces stratégies classificatoires permettent une immense économie de temps dans le parcours exploratoire d'un corpus, et à ce titre, elles sont incontournables lorsqu'on est confronté à de vastes corpus textuels. D'autre part, elles servent d'indices pour détecter rapidement certains liens sémantiques et textuels. Cependant, lorsqu'associées à des stratégies linguistiques plus fines et intégrées dans des systèmes hybrides (i.e., avec analyseurs linguistiques d'appoint), elles livrent une assistance précieuse pour des analyses globales. Elles permettent un premier déblaiement général du texte. Peuvent alors suivre des analyses plus fines.

Les recherches récentes permettent de penser qu'on peut améliorer ces techniques de classification de l'information. En effet, de nouveaux modèles classifieurs dits émergentistes commencent à être explorés pour ce type de tâche. Ils ont pour fondement théorique que le traitement «intelligent» de l'information est avant tout associatif et surtout adaptatif. Parmi ces modèles dits «de computation émergente» on distingue les modèles «génétiques», markoviens et surtout connexionnistes. Parmi ces derniers, on trouve une grande variété de modèles, entre autres, les modèles matriciels linéaires et non linéaires, les modèles thermodynamiques, et les modèles basés tantôt sur la compétition, tantôt sur la rétropropagation, mais surtout sur des règles complexes d'activation et d'apprentissage. Les principaux avantages de ces modèles tiennent au fait que leur structure parallèle leur permet de satisfaire un ensemble de contraintes qui peuvent être faibles et même, dans certains cas, contradictoires et de généraliser leur comportement à des situations nouvelles (le filtrage), de détecter des régularités et ce, même en présence de bruit. Outre les propriétés de généralisation et de robustesse, la possibilité pour ces modèles de répondre par un état stable à un ensemble d'inputs variables repose sur une capacité interne de classification de l'information.

Cependant, tous ces modèles classifieurs émergentistes opèrent souvent sur des données bien contrôlées et qui toutes doivent être présentes au début et tout au long du traitement. De plus, ils exigent souvent divers paramètres d'ajustement qui relèvent souvent d'une description statistique du domaine. Il s'en suit que les résultats de classification obtenus sont valides pour autant qu'ils portent sur les données bien contrôlées où peu de modifications sont possibles. Si, après la période d'apprentissage, pour quelque raison que ce soit, les systèmes sont confrontés à des données qui n'étaient pas prévues dans les données de départ, ils auront tendance à les classer dans les prototypes déjà construits, donc à produire une sous-classification.

Or, dans le domaine du texte, nous sommes confrontés à des corpus en constante modification. Chaque nouvelle page peut possiblement contenir des informations que le système peut ne jamais avoir rencontrées, et donc qu'il ne peut se permettre de classer dans ses prototypes antérieurement construits. Il faut donc, outre la dynamique (incrémentalité) de l'apprentissage, un système qui soit aussi plastique, c'est-à-dire qui s'adapte à de

nouvelles données. Et on voit apparaître depuis quelque temps des recherches qui sont de plus en plus sensibles à cette dimension (Burr, 1987; Veronis et al., 1990; Balpe et al., 1996, etc.). Et c'est dans cette perspective que la présente recherche a été effectuée. Nous en présentons ici les résultats préliminaires à l'occasion d'une application d'analyse terminologique.

4. LA MÉTHODE

Dans sa forme concrète et expérimentale, la recherche a consisté à explorer un modèle connexionniste pour extraire de l'information de type terminologique sur des fichiers textuels. La réalisation de cette recherche comporte les étapes suivantes :

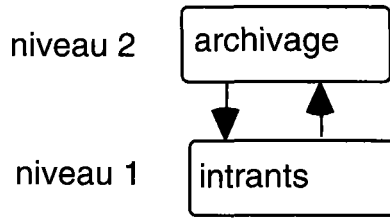
- 1- une modélisation d'un traitement connexionniste du texte;
- 2- une expérimentation d'une chaîne de traitement sur des textes.

4.1 Modélisation d'un traitement connexionniste de textes

Il n'existe pas, à notre connaissance, un grand nombre de modèles connexionnistes pouvant confronter simultanément les problèmes de dynamicité et de plasticité de l'information. La dynamicité est la capacité du système à traiter de manière adaptative les informations qu'il reçoit. Quant à la plasticité, il s'agit de la capacité du système à traiter de l'information pour laquelle il n'avait pas été paramétrisé. Un des modèles qui avait ces visées est ART 1 (Grossberg et Carpenter, 1987). En effet, dans sa définition originale le modèle ART 1 se veut un système classifieur auto-organisationnel, sans supervision, pouvant opérer sur des stimuli non contrôlés et bruités.

Ce modèle a évolué à travers les années. Il a donné naissance à de nombreuses variantes de la règle de transmission et de la règle d'apprentissage. Il a introduit des facteurs de multiplication dans l'activation et des facteurs de dégradation dans l'encodage de l'information. L'une de ses prétentions importantes est qu'il est en mesure de traiter de manière adaptative des stimuli qui sont changeants (plasticité), c'est-à-dire qui ne font pas partie d'un corpus contrôlé d'avance. L'objectif ultime de ce modèle est de créer une grande stabilisation dans la représentation des patrons de stimuli. Au début, ART 1 ne pouvait traiter que des informations de nature binaires. Plus tard, le modèle ART 2 accepte des informations dont les valeurs ne sont plus discrètes ou binaires. Enfin dans ART 3 le modèle a consolidé ses stratégies et offre un traitement plus fiable.

L'idée centrale du modèle ART est celle d'un système d'interaction entre deux niveaux qui entrent en résonance mutuelle.



Le système reçoit dans un premier niveau 1 des stimuli qui sont envoyés, mais aussi modifiés (selon une distribution et un poids particulier) au deuxième niveau 2, qui est un niveau d'archivage.

Arrive donc au deuxième niveau un pattern différent de ce qui était à l'intrant. Il y a alors comparaison par un processus dit de résonance. Si le nouveau pattern n'a aucune ressemblance avec les anciens, il sera alors conservé et servira de gabarit ou de prototype avec lequel les intrants nouveaux seront éventuellement comparés. En fait, le pattern au niveau 2 servira de modèle de comparaison avec les nouveaux intrants. S'il diffère de ces derniers, un autre pattern sera essayé jusqu'à ce qu'une correspondance soit acceptable (selon certains paramètres) et s'il y a correspondance acceptable l'intrant sera alors classé avec le prototype. Mais s'il n'est pas acceptable, le nouveau pattern sera considéré comme un prototype en émergence et il servira éventuellement de nouveau gabarit aux autres intrants que le système rencontrera.

La correspondance entre le pattern prototypal et le pattern intrant est la résonance. Au fur et à mesure que l'apprentissage se poursuit, il y a consolidation de cette résonance. L'adaptabilité survient par la modification constante des interconnexions entre les niveaux. Pour réaliser cette interaction le système doit être contrôlé par divers paramètres qui assurent la solidité du traitement.

4.2 L'application du modèle ART

La deuxième étape de la recherche est une expérimentation de ce modèle sur des textes et sur une tâche spécifique. Un texte regorge de connaissances de plusieurs types qui peuvent en être extraites. On ne lit pas un texte uniquement pour savoir qui a fait quoi (connaissances du monde). On le lit pour connaître des faits, certes, mais aussi des actions, des valeurs, des jugements etc. (Meunier, 1996). Une des connaissances qu'on cherche est aussi de nature métalinguistique, i.e. qui porte sur la nature de la langue même du texte, par exemple, son style, sa terminologie, son lexique, etc. Dans la présente recherche, nous avons choisi d'explorer l'extraction des connaissances métalinguistiques de types terminologique et sémantique. Pour ce faire nous avons développé une *chaîne de traitement* modulaire qui tente d'intégrer les étapes du travail analytique du terminologue ou du documentaliste devant travailler à l'identification des champs lexicaux d'un terme particulier. Par exemple, la chaîne de traitement pourrait aider le terminologue à identifier rapidement que, dans un texte, le mot CODE peut avoir des champs sémantiques différents où ce mot prend le sens de CODE civil, CODE de la route, CODE informatique, CODE de la construction, etc.

Cette chaîne de traitement a été appliquée à deux textes spécifiques qui nous servent ici d'illustration : le premier de 900 pages, la *Convention de la Baie James* d'Hydro Québec; le second, la revue *Spirale* (Belgique) de quelque 180 pages.

L'expérimentation sur les textes

Dans la première étape de sa gestion, le texte est reçu et traité par des modules d'analyse de la plate-forme ALADIN-TEXTE. Cette plate-forme est un atelier qui utilise des modules spécialisés dans l'analyse d'un texte. Dans un premier temps, un filtrage sur le lexique du texte est fait. Par divers critères de discrimination, on élimine du texte certains mots accessoires (mots fonctionnels ou statistiquement insignifiants, etc.) ou ceux qui ne sont pas porteurs de sens d'un point de vue strictement sémantique, et dont la présence pourrait nuire au processus de catégorisation, soit parce qu'ils alourdiraient indûment la représentation matricielle, soit parce que leur présence nuirait au processus interprétatif qui suit la tâche de catégorisation. Vient ensuite une description morphologique minimale de type lemmatisation. Cette opération consiste à remplacer chaque mot par son équivalent canonique. (e.g. aimerions --> AIMER). Ce processus se justifie par le fait que les déclinaisons propres à la grammaire ou à la syntaxe d'une langue n'affectent en rien le contenu sémantique réel des termes. De la même façon, remplacer un mot décliné (soit dans sa forme verbale, adverbiale, adjectivale, pronominale ou autres) par sa forme nominale n'a aucun impact significatif sur le contenu sémantique principal de ce dernier. Ces dimensions morphologiques touchent surtout des modalités tels : le genre, l'aspect, le temps, etc.

Puis une transformation est opérée pour obtenir une représentation matricielle du texte. Cette transformation est encore effectuée par des modules d'ALADIN explicitement dédiés à cette fin. On produit ainsi un fichier indiquant pour tout lemme choisi sa fréquence dans chaque segment du texte. Suit ensuite un post-traitement pour construire une matrice dans un format acceptable par les réseaux de neurones. Dans la présente expérimentation, la précédente matrice est alors soumise aux classifieurs ART. Selon le réseau utilisé (ART 1 ou FUZZY-ART), la matrice générée peut être constituée exclusivement de données binaires (ART 1) ou de données non binaires (FUZZY-ART). Dans le cas du réseau ART 1, les données subissent alors une réduction, puisque la fréquence d'apparition des lemmes est alors remplacée par une simple indication de présence ou d'absence.

Les réseaux de type ART se prêtent particulièrement à ces contraintes. Le Fuzzy-Art semble donner des résultats supérieurs car le processus de catégorisation produit moins de classes contenant un seul élément caractéristique (un seul lemme), ce qui du point de vue du terminologue n'est d'aucune utilité.

Les résultats de la classification neuronale

Appliqué à la matrice précédente, ART génère des classes de SEGMENTS qui présentent entre eux une certaine similarité lexicale. Autrement dit, chaque classe de segments constitue pour ART un «prototypé» qui est donc caractérisée par les termes qui sont présents également dans tous les segments du texte. Une fois trouvées ces classes de segments présentant une similarité lexicale, on en extrait, pour chacune, le lexique, c'est-à-dire on trouve pour les termes qui la caractérise.

Ensuite, on choisit un terme particulier, et on étudie les classes dans lesquelles il apparaît. C'est ainsi, par exemple, que dans l'analyse sur la revue *Spirale*, nous avons obtenu la classification suivante pour les termes *rapport* et *tête*

Le terme *rapport*

apparaît dans les classes 28 35, 39 40 54,

la classe 28 : les segments 71-73

contient : *choix connaissance, document, façon fiction personnage, rapport, savoir et travail*

la classe 35, les segments 89-92 qui contiennent les mots

autres connaissances doute formes image processus production et rapport

la classe 39 : 100 101 qui contiennent les mots

autres élèves enseignant ensemble genre, jeunesse rapport roman

la classe 40 : 102 103 qui contiennent les mots

écrit, écriture élémentaires, jeunesse, monde problème rapport réel, scolaire, situation

la classe 54 : 59 64 qui contiennent les mots

auteurs, autres, discours, jeunesse, lecture, mode, question, rapport, rôle, temps, vie

Sur la distribution de ce terme dans les classes de segments trouvées, on peut tenter l'interprétation sémantique suivante. On peut dire que le terme *rapport* est utilisé dans deux ensembles de contextes relativement différents.

Un premier pointe vers le concept de *rapport* comme *document* où est déposé de l'information (classe 28 dans les segments 71-73).

Un deuxième pointe vers le concept des *liens entre des individus et autres chose* (39, 40, 54).

Enfin la classe 35 n'est pas clairement situable dans l'un ou l'autre sens précédent.

Cependant, on voit bien que dans ce texte, ces deux significations sont les deux seules possibles. Pour un terminologue, le terme *rapport* dans la revue *Spirale* # 15 n'est pas employé dans le sens suivants : d'une proportion, c'est-à-dire d'un rapport logique, d'un rapport financier, d'une maison de rapport, d'une communication ou encore d'une perspective, etc.

Une analyse similaire peut être tentée sur le terme *tête*.

Nous retrouvons le terme *tête* dans les classes dans la classe 20 et la classe 58.

La classe 20 : les segments 53 et 107.

façon, moment, place, rôle, savoir, tête.

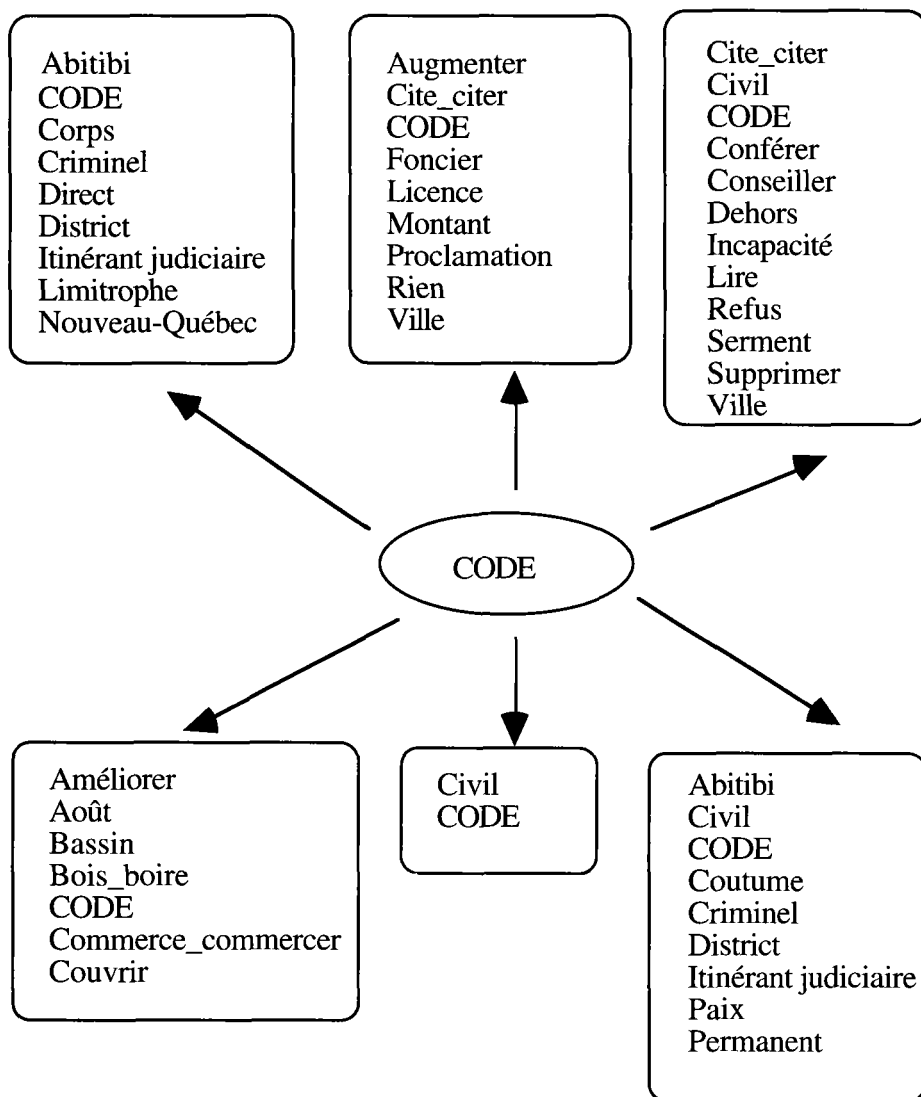
La classe 58 : le segment 93

aide, bon, comprendre, connaissance, École, ensemble, histoires, jour, lecteur, loisirs, narratif, père, perspective, schéma, situation, souris, suite, tête, textuels, titre, traitements.

Ces deux classes montrent que le terme *tête* est utilisé dans deux sens différents. Pour la classe 20 le sens du terme tête est proche de la signification «leader» or dans la classe 58 la signification est plus proche de l'interprétation «intelligence».

Le post-traitement

La présentation linéaire de ces résultats rend l'interprétation difficile à réaliser. On peut imaginer une meilleure convivialité de la configuration des résultats. En effet, une légère transformation peut leur donner une lecture plus facile. À titre d'exemple nous présentons ci-après la configuration d'un terme plus riche d'acception sémantique que nous avons trouvé dans le texte d'Hydro Québec : la *Convention de la Baie James* (870 p.). Dans ce texte, le mot CODE se retrouve dans plusieurs classes. Il est alors possible pour ce lexème de dessiner le graphe des classes dans lesquelles il apparaît.



On voit ainsi apparaître la différence des réseaux sémantiques de ce terme, qui tantôt est utilisé dans le sens de code civil, tantôt dans le sens de code de comportement, tantôt dans le sens du code criminel, etc. Pour un terminologue, ce graphe est intéressant. Il sert d'indice au réseau sémantique de ce terme, et donc de ses acceptions dans le discours particulier. La thèse associative classique explique ce fait en postulant que si deux termes se retrouvent ensemble dans un même contexte, c'est que leurs contenus sémantiques ou conceptuels sont associés.

5. INTERPRÉTATION ET VALIDATION DES RÉSULTATS

Les résultats produits sont alors prêts à être validés. Une méthode de validation inter-juges a été utilisée pour évaluer la qualité et la pertinence de la classification opérée par le réseau de neurones. La méthode consiste essentiellement à comparer les évaluations faites et les appréciations effectuées par les différentes personnes impliquées dans l'évaluation des résultats. Le travail a consisté essentiellement en une analyse des classes produites pour déterminer leurs pertinences d'un point de vue terminologique. De façon générale, l'analyse a montré que plus de 80 % des classes constituées seraient pertinentes et utilisables dans le processus de création des fiches terminologiques. Mais un travail de validation sur un corpus banc d'essai est maintenant nécessaire.

6. CONCLUSION

L'objectif de notre recherche était de permettre l'extraction de connaissances terminologiques à partir du texte plein. Cette extraction devait, de plus, pouvoir se faire sur un corpus en évolution constante (plasticité) et les catégorisations effectuées se devaient de rester pertinentes et utilisables. Le processus de catégorisation quel qu'il soit se devait d'opérer sans supervision aucune (adaptabilité) et sans faire appel à des connaissances préformées ou prédigérées, celles-ci n'étant simplement pas disponibles dans le cas qui nous concerne.

Une méthode connexionniste comme solution au problème de l'extraction terminologique sur des textes entiers a été expérimentée avec des résultats très encourageants. L'approche choisie montre un intérêt certain et des avantages indéniables. Par exemple, le gain de temps estimé par rapport au travail manuel requis pour effectuer un travail terminologique équivalent est considérable. De plus, la précision et la richesse des suggestions faites par le système ne sont en aucune mesure comparables avec ce qu'on obtient avec les méthodes actuelles.

L'approche choisie s'inscrit parfaitement dans l'optique d'une solution opérationnelle aux problèmes flagrants et réels qui minent l'industrie de la langue et toute industrie qui implique la manipulation et la classification de masses importantes de documents.

Plusieurs variantes de l'expérimentation sont possibles et envisagées. On pourra opérer les classes autrement que par l'intersection entre les segments. Par exemple, tirer des informations pertinentes de l'union des unifs présents dans les segments regroupés. On pourrait sûrement obtenir des résultats encore plus probants en tenant compte de la dépendance entre les variables (ou unifs) pris en considération. Un pré-processing sémantique (non approfondie pour conserver l'avantage de temps) pourrait amener des améliorations considérables. Plusieurs variantes des filtres de pré-traitement utilisés sont à l'étude. Le filtrage à la sortie des classes ne représentant pas une richesse d'information suffisante (selon un critère donné, ex. : pas assez d'unifs dans cette classe) améliorerait les résultats.

Plusieurs problèmes restent à résoudre (grandeurs fixes des intrants, dégradation du temps d'apprentissage avec le nombre des intrants, interprétation, etc.).

Un module d'interprétation des résultats (avec interface graphique) est en cours de développement.

RÉFÉRENCES

- BALPE, J. P., LELU, A., PAPY, F., & I. S. (1996) : *Techniques avancées pour l'hypertexte*, Paris, Hermes.
- BURR, D. J. (1987) : «Experiments with a Connectionist Text Reader», *IEEE First International Conference on Neural Networks*, San Diego, pp. 717-724.
- CARPENTER, G. & G. GROSSBERG (1991) : «An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition», *Neural Networks*, 4, pp. 493-504.
- CHURCH, K., GALE, W., HANKS, P. & D. HINDLE (1989) : «Word Associations and Typical Predicate-argument Relations», *International Workshop on Parsing Technologies*, Carnegie Mellon University, Aug. 28-31.
- CHURCH, K. W. & P. HANKS (1990) : «Word Association Norms, Mutual Information, and Lexicography», *Computational Linguistics*, 16, pp. 22-29.
- DELISLE, S. (1994) : *Text Processing Without A Priori Domain Knowledge : Semi Automatic Linguistic Analysis for Incremental Knowledge Acquisition*, PhD Thesis, Ottawa University.
- GARNHAM, A. (1981) : «Mental Models and Representation of Texts», *Memory and Cognition*, 9, pp. 560-565.
- GREFENSTETTE, G. (1992) : «Sextant : Exploring Unexplored Contexts for Semantic Extraction from Syntactic Analysis», *Proc. of the 30th Annual Meeting of the ACL*, pp. 324-326.
- GREFENSTETTE, G. (1992) : «Use of Syntactic Context to Produce Term Association Lists for Text Retrieval», *Proc. of SIGIR 92 ACM*, Copenhagen, June 21-24.
- GROSSBERG, S. & S. CARPENTER (1987) : «Self Organization of Stable Category Recognition Codes for Analog Input Patterns», *Applied Optics*, 26, pp. 4919-4930.
- JACOBS, P. & U. ZERNIK (1988) : «Acquiring Lexical Knowledge from Text A Case Study», *Proceedings of AAAI 88*, St Paul, Min.
- KOHONEN, T. (1982) : «Clustering, taxonomy and topological Maps of Patterns», *IEEE Sixth International Conference on Pattern Recognition*, pp. 114-122.
- LEBART, L. & A. SALEM (1988) : *Analyse statistique des données textuelles*, Paris, Dunod.
- LELU, A. (1995) : «Hypertextes : la voie de l'analyse des données», L. Bolasco, S. L. A. Salem (Eds), *Anilisi statistica dei dati testuali* , vol. 2, Rome, CISU, pp. 85-96.
- LIN, X., SOERGEL, D. & G. MARCHIONINI (1991) : «A Self Organizing Semantic Map for Information Retrieval», *SIGIR 91*, Chicago, Illinois.

- MEUNIER, J.-G. (1996) : «Théorie cognitive : son impact sur le traitement de l'information textuelle», V. Riale et D. Fisette, *Penser l'esprit des sciences de la cognition à une philosophie cognitive*, Presses de Université de Grenoble, pp. 289-305.
- MOULIN, B. & D. ROUSSEAU (1990) : «Un outil pour l'acquisition des connaissances à partir de textes prescriptifs», *ICO*, Québec 3 (2), pp. 108-120.
- REGOCZEI, S. et al. (1988) : «Creating the Domain of Discourse : Ontology and Inventory». Gaines & Boose (Eds), *Knowledge Acquisition Tools for Experts and Novices*, Academic Press.
- REGOCZEI, S. & G. HIRST (1989) : On extracting knowledge from Text. Modeling the Architecture of Language Users. (TR CSRI 225), Computer Systems Research Institute, University of Toronto.
- SALTON, G. (1988) : «On the Use of Spreading Activation», *Communications of the ACM*, vol 31 (2).
- SALTON, G., ALLAN, J. & C. BUCKLEY (1994) : «Automatic Structuring and Retrieval of Large Text File», *Communications of the ACM*, 37 (2), pp. 97-107.
- SEFFAH, A. & J.-G. MEUNIER (1995) : «ALADIN : un atelier orienté objet pour l'analyse et la lecture de textes assistée par ordinateur», *International Conference on Statistics and Texts*, Rome.
- TAPIERO, I. (1993) : *Traitement cognitif du texte narratif et expositif et connexionnisme : expérimentations et simulations*, Université de Paris VIII.
- THRANE, T. (1992) : «Dynamic Text Comprehension», J. O. S. Jansen, H. Prebensen, T. Thrane (Eds), Copenhaguen, Museum Tuscalanum Press.
- VERONNIS, J., IDE, N. M. & S. HARIE (1990) : «Utilisation de grands réseaux de neurones comme modèles de représentations sémantiques», *Neuronimes*.
- VIRBEL, J. (1987) : «L'apport de connaissances linguistiques à l'interprétation des structures textuelles», *Structure des documents, Bigre++Globule*, 53, pp. 77-97.
- VIRBEL, J. E., F. PASCUAL (1992) : *La lecture assistée par ordinateur, Rapport de recherche*, Toulouse, Laboratoire IRIT.
- VIRBEL, J. (1993) : «Reading and Managing Texts on the Bibliothèque de France Stations», P. Delany & P. Landow (Eds), *The Digital Word : Text Based Computing in the Humanities*, Cambridge, Mass., MIT Press.
- YOUNG, T. & T. CALVERT (1987) : *Classification, Estimation, and Pattern Recognition*, Amsterdam, Elsevier.
- ZARRI, G. P. (1990). «Représentation des connaissances pour effectuer des traitements inférentiels complexes sur des documents en langage naturel», Office de la langue française (Ed.), *Les industries de la langue. Perspectives 1990*, Gouvernement du Québec.

VERS LA SPÉCIFICATION ET L'EXTENSION DES RELATIONS TERMINOLOGIQUES : TYPOLOGIE ET INSERTION DANS UN SYSTÈME DE SIGNIFICATIONS DES RELATIONS¹

Christophe JOUIS^(1et2), Widad MUSTAFA ELHADÍ⁽¹⁾, Ismaïl BISKRI^(2et3), Florence LE PRIOL⁽²⁾

(1) *IDIST - Information, Documentation, Information Scientifique et Technique, Université Charles De Gaulle-Lille III, Villeneuve d'Ascq, France*; (2) *LALIC-CAMS, Université de la Sorbonne-Paris IV, Paris, France*; (3) *LANCI, Université du Québec à Montréal, Canada*,

1. INTRODUCTION

Les spécifications des relations terminologiques (c'est-à-dire, pour chaque relation, ses propriétés algébriques, le nombre et les types de ses arguments, etc.) posent des problèmes particuliers résultant du fait que les principaux utilisateurs de ces relations sont des acteurs du monde industriels. «Bien qu'issues de concepts relevant de la linguistique et de concepts propres à la terminologie, elles ont été élaborées dans le cadre d'actions de normalisation industrielle et de ce fait opèrent des raccourcis difficilement acceptables dans un contexte scientifique» (Humbley 1997). Les liens notionnels se réduisent à la distinction établie par les normes ISO 704 (1987) et ISO 1087 (1990) entre les relations hiérarchiques (relations espèce-genre et partie-tout), et les relations non hiérarchiques («rapports de contiguïté dans le temps, dans l'espace, de connexion causale, etc.»). Une des

¹ Ces réflexions sont en partie issues des travaux que nous menons actuellement dans le cadre d'une d'une Action de Recherche Partagée (ARP) soutenue par l'Agence francophone pour l'enseignement supérieur et la recherche (AUPELF-UREF) : «Intégration d'une approche sémantique linguistique et d'une approche numérique pour un outil d'aide à la construction de bases terminologiques». Les composantes de notre ARP sont les équipes IDIST/CREDO, LALIC-CAMS, et LANCI. L'expérience consiste à intégrer dans une même chaîne de traitement deux stratégies complémentaires : (1) une analyse connexionniste (le système CONTERM, conçu par le LANCI) avec (2) une analyse sémantique linguistique (le système SEEK, conçu par l'équipe LALIC et l'IDIST). L'objectif de ce projet est la réalisation d'un outil d'aide à la construction de terminologies à partir de documents scientifiques ou techniques. Pour ce faire, nous procédons en deux étapes successives. La première phase est effectuée par le système CONTERM. Elle consiste en un filtrage du corpus grâce à une analyse numérique qui fournit une classification et une structuration du corpus en des classes de termes. La deuxième phase est effectuée par le système SEEK. Elle consiste à identifier des relations sémantiques entre les termes des classes obtenues par CONTERM. (Jouis et al. 1997). Dans cet article, nous présentons l'ensemble structuré des relations sémantiques utilisées dans ce projet pour relier les termes entre eux.

approches possibles pour aborder ce problème consiste à organiser les relations dans une typologie. Par exemple, Winston et al. (1987) distinguent différents types de relations partie-tout. Cette typologie a par exemple inspiré le traitement des relations partie-tout dans le dictionnaire électronique WordNet (Miller 1990). Divers travaux récents d'application de la terminologie à l'accès à l'information, notamment la construction de thesaurus, tentent de mieux spécifier les propriétés des liens entre concepts et étendre les relations non hiérarchiques (cf. par exemple Molholt 1996; Green 1996; Bean 1996; etc.).

D'autre part, de travaux récents visent à intégrer dans les modèles terminologiques des théories issues de la linguistique (sémantique linguistique, par exemple) et de l'Intelligence Artificielle (modélisation des connaissances pour la conception de systèmes à base de connaissances). Pour toutes ces disciplines, le problème fondamental est la nécessité de structurer la connaissance dans un micro-domaine :

En Intelligence Artificielle, des méthodes d'acquisition et de modélisation des connaissances telles que KADS II (Wielingua et al. 1992)² ou METHODAC (Jouis 1993) ont été développées pour l'aide à la conception de systèmes à base de connaissances. Ces méthodes proposent des schémas de structuration des connaissances pour modéliser un domaine d'expertise sous forme de concepts reliés par des relations sémantiques dans des langages à objets (appelé «niveau domaine dans KADS» et «couche descriptive» dans METHODAC). Ces langages nous semblent très proches des structures des bases de données terminologiques.

En terminologie, parallèlement, plusieurs équipes ont développé des logiciels pour «naviguer» dans des réseaux de concepts structurant des micro-domaines. Citons en particulier le système Termisti (Van Campenhoudt 1994), le système Code : projet COGNITERM (Meyer et McHaffie 1994) ou le système Ikarus (Université d'Ottawa³) qui permettent la gestion de bases de connaissances terminologiques (B.C.T.).

Aussi, dans la même perspective de mieux structurer la terminologie d'un domaine de connaissances, et plus particulièrement pour l'aide à l'indexation de documents et/ou la recherche d'informations, nous proposons de nous appuyer sur un ensemble de relations issues d'un modèle linguistique. Il s'agit de la Grammaire Applicative et Cognitive (GAC, Desclés 1990) et de ses extensions pour l'acquisition et la modélisation des connaissances (Jouis 1993). Les relations de ce modèle s'insèrent dans un système de significations de repérage entre entités. Le repérage se spécifie suivant les propriétés algébriques qui lui sont attribuées axiomatiquement en diverses relations. À la différence de la typologie établie par Felbert (1987) fondée sur la distinction entre les «rapports logiques», les «rapports ontologiques» et les «rapports d'effet», notre typologie s'appuie essentiellement sur la distinction entre situation statique (état de chose dans le domaine) et situation dynamique (modification et changement dans le domaine).

Dans la deuxième partie, nous présenterons l'architecture et les quatre catégories primitives sémantiques du modèle linguistique de la GAC.

² Associée aux logiciels Open KADS (Bull) et ILOG KADS Tool (ILOG).

³ Cf : Meyer et Skuce (1997) : «Bases de connaissances et bases textuelles sur le Web : le système Ikarus», V^{es} Journées scientifiques : La mémoire des mots, Tunis, 25, 26, 27 septembre 1997.

Puis, dans la troisième partie, nous présenterons un sous-ensemble des relations sémantiques de la GAC que nous utilisons actuellement dans trois projets d'aide à la construction de terminologies⁴.

2. LA GAC ET LES PRIMITIVES SÉMANTIQUES

2.1 L'architecture de la GAC

La Grammaire Applicative et Cognitive (GAC) est organisée comme un système de changement de représentations. En particulier, ce modèle distingue plusieurs niveaux de description : un niveau phénotype, un niveau génotype et un niveau cognitif. Le phénotype correspond au niveau morpho-syntaxique d'une langue particulière (lexique, syntaxe, règles morphologiques, etc.). La description de ce niveau est effectuée dans le cadre d'une grammaire catégorielle étendue (Biskri 1995). Le niveau génotype correspond aux structures grammaticales et sémantiques plus générales du langage. En particulier, on construit à ce niveau les représentations des relations prédicatives. Une relation prédicative est constituée d'une combinaison d'opérateurs grammaticaux (voix passive/active, temps et aspect, détermination, etc.) qui s'applique à un prédicat et ses arguments. Au niveau génotype, les prédicats sont vus comme des fonctions à valeur dans {vrai, faux}. Lorsque l'on passe au niveau cognitif, on décrit une représentation du sens des prédicats et des arguments sous forme d'une combinaison de primitives. Par exemple, l'énoncé :

Jean entra dans la pièce

est vu tout d'abord comme un événement (EVE) passé (PASSE) représenté sous la forme d'une expression applicative :

PASSE (EVE ([entrer_dans] [la_pièce] [Jean])).

Au niveau cognitif, il s'agit ensuite de donner la représentation du prédicat [entrer_dans] sous forme d'une combinaison d'opérateurs élémentaires et de typer les arguments [la_pièce] et [Jean]. Les primitives sémantiques proposées dans le modèle de la GAC sont en partie fondées sur les représentations cognitives construites par la perception visuelle de l'espace (positions d'objets par rapport à des lieux), la perception des stabilisations, de changements perçus dans le temps et de l'action. Une catégorisation s'effectue donc autour de l'opposition stable / non stable, c'est-à-dire statique // cinématique/dynamique.

⁴ Les trois projets sont les suivants : (1) le projet ARP de l'AUFELF-UREF déjà mentionné en note; (2) un projet soutenu par l'AUFELF-UREF dans le cadre d'une Action de Recherche Concertée : «Évaluation d'outils d'aide à la construction de terminologie et de relations sémantiques entre termes à partir de corpus», organisée par l'équipe IDIST/CREDO (Beguin et al. 1997) et (3) un projet qui vient d'être retenu par le Ministère de l'Éducation Nationale, de l'Enseignement supérieur et de la Recherche français (section DISTB) & le Ministère des Postes et des Télécommunications français «Intégration d'une approche sémantique linguistique et d'une approche numérique pour un outil de filtrage d'informations sur les réseaux», équipes IDIST/CREDO & le TIMC-IMAG (Techniques de l'Imagerie, de la Modélisation et de la Cognition), UMR CNRS 5525, Université Joseph Fourier (Grenoble).

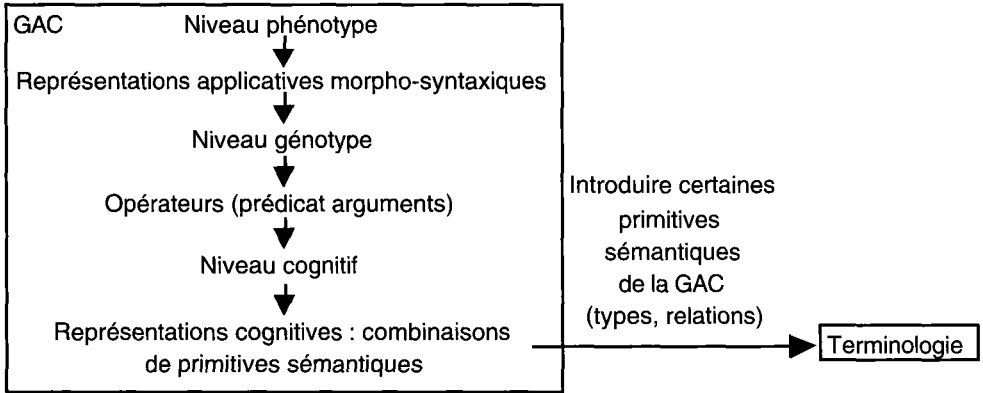


Figure 1 : Architecture de la GAC et introduction de primitives dans la terminologie

Notre objectif consiste alors à introduire certaines primitives sémantiques de la GAC pour mieux spécifier les relations terminologiques.

2.2 Les primitives

Au niveau cognitif, nous distinguons quatre catégories de primitives :

1. des **types sémantiques** élémentaires permettant de classer les notions manipulées (entités individualisables, entités booléennes, entités massives, classes collectives, classes distributives, lieux, etc.);
2. des opérateurs formateurs de **types complexes** à partir des types de base (produits cartésiens, types fonctionnels, etc.);
3. des **relations statiques** fondamentales (Desclés 1987; Jouis 1993) entre unités terminologiques. Elles permettent de décrire des états d'un domaine de savoir. Nous distinguons en particulier les identifications (identité, égalité), incompatibilités, mesures, cardinalités, comparaisons, inclusions (entre classes distributives), relations d'appartenance (d'une entité individuelle à une classe distributive, objet typique/atypique), localisations d'une entité par rapport à un lieu (intérieur, extérieur, frontière, fermeture, orientation, etc.), relations partie/tout (ingrédients directes, nécessaires, atomiques, typiques, uniques, quantifiables, etc.);
4. des **relations dynamiques** fondamentales entre unités terminologiques. Par opposition aux situations statiques, ces relations permettent de décrire des processus ou des événements d'un domaine de savoir (unités terminologiques caractérisant des situations cinématiques ou dynamiques) : mouvements spatio-temporels (orientés, non orientés), changements d'états, conservations d'un mouvement, itérations, variations d'intensité, les capacités d'effectuer une action, de la contrôler, de viser un but, de planifier une action et les relations de causalité.

Les primitives précédentes se combinent entre elles pour former des schémas de représentation. Par exemple, le prédicat [entrer_dans] appliqué à ses arguments dans l'expression applicative :

PASSE (EVE ([entrer_dans] [la_pièce] [Jean])).

sera représenté au niveau cognitif par la combinaison de primitives suivante :

[entrer_dans] = **CONTR** [Jean] (**MOUV** (loc-ex [Jean] [la-pièce]) (loc-in [Jean] [la-pièce]))

Ce qui signifie qu'il y a un mouvement (MOUV) contrôlé (CONTR) par Jean, qui fait passer d'une situation statique dans laquelle Jean est à l'extérieur de la pièce à une nouvelle situation statique où Jean est à l'intérieur de la pièce.

2.3 Le système de types

Toutes les représentations sont typées soit par un type élémentaire, soit par un type complexe. Les types élémentaires sont définis de la manière suivante :

- Les entités **booléennes** (notées H) sont des objets qui ont pour valeur vrai ou faux.
- Les entités **individualisables** (notées J) sont celles que l'on peut désigner et montrer par pointage. On peut les compter individuellement ou les regrouper en classes. Un opérateur de quantification permet de construire une classe plus ou moins déterminée d'exemplaires individualisés. Par exemple, les entités *Jean*, *table*, *chaise*, *meuble*, *homme*, *enfant* sont individualisables.
- Les entités **massives** (notées M) telles que *eau*, *mer*, *sable*, *vin*, *beurre*, *blé* ne sont pas des entités individualisables. Notons cependant qu'un certain nombre d'opérateurs (des classificateurs) permettent de rendre individualisable une notion massive : *un verre d'eau*, *un pâté de sable*, *un bras de mer*, *une bouteille de vin*, *un morceau de beurre*.
- Les **classes distributives** (notées D) rassemblent des entités individuelles ayant une même propriété. Par exemple, *être-un-carré* représente une classe d'individus (ou «concept»).
- Les **classes collectives** (notées C) se distinguent des classes distributives parce qu'elles représentent des objets qui forment un «tout» à partir d'objets plus élémentaires. Cette différence a été mise en évidence par Lesniewski⁵. Les classes distributives associées

⁵ Lesniewski (1886-1939) a proposé une théorie générale des tous et des parties (la méréologie), en réponse aux problèmes des paradoxes de la théorie des ensembles (Cantor, Russel & WhiteHead, Frege). Une analyse détaillée de la méréologie a été réalisée par (Mieville 1984). Considérant une classe comme un objet, un tout, vu comme le «tas» d'éléments qui le constituent, disjoints ou non. Lesniewski aboutit à la conclusion que la notion de classe contient deux caractères : le distributif et le collectif. L'exemple suivant, emprunté à Grize (1973 : 86), donne une idée de la différence :

«Une classe distributive est, à proprement parler, l'extension d'un concept. Si p est le concept planète, dire que Jupiter est une planète, c'est soit poser pJ, soit $J \hat{=} \{x / px\}$ et l'information transmise est la même dans les deux écritures. Soit donc la classe distributive :

p = {Mercure, Vénus, la Terre, Mars, Jupiter, Saturne, Uranus, Neptune, Pluton}.

aux relations d'appartenance et d'inclusion, tandis que les classes collectives sont associées aux relations partie-tout. Ainsi, *foule, armée, armada, famille* sont des classes collectives.

Le type des **lieux** (notés L) représente des étendues ou des regroupements de positions d'une même entité (individualisable, collective ou massive) : *Paris, jardin, maison*.

Nous remarquons qu'une même unité linguistique peut se voir attribuer plusieurs types suivant le point de vue. Par exemple, *Paris* peut être vue comme un lieu (*Jean est à Paris*), comme une classe collective (*Paris est composée de 20 arrondissements*), comme une entité individuelle (*Paris est une capitale*), etc. (Jouis 1993).

À partir de l'ensemble S des types élémentaires de la GAC, on peut alors définir un système de types dérivés de façon récursive à partir des règles suivantes :

- (i) les éléments de S sont des types (élémentaires);
- (ii) Si x et y sont des types, alors F_{xy} est un type (fonctionnel)⁶.

Le symbole **F** est un opérateur formateur de types fonctionnels. Une entité f de type F_{xy} (notée $F_{xy} : f$) est un opérateur unaire qui prend pour argument un objet de type x pour fournir un résultat de type y. Considérons une entité a de type x, l'application de f sur a construira une certaine entité b de type y :

$(F_{xy} : f \ x : a) \ y : b$.

Par exemple, le type FJH est celui d'un opérateur qui s'applique à une entité individualisable (J) pour retourner une valeur de vérité H (propriété unaire des individus, classe d'individus ou «concept», tel que «être-un-carré»). Une relation entre une entité individuelle et un lieu (localisation) aura le type FJFLH⁷. Le connecteur logique «ET» est un opérateur binaire qui s'applique à deux entités booléennes : FHFHH etc. De cette manière, toutes les représentations du niveau cognitif sont typées.

Elle contient neuf éléments et rien d'autre. Les calottes polaires de Mars, la tache rouge de Jupiter, les anneaux de Saturne n'appartiennent pas à p. Et pourtant tout cela et mille choses ont affaire avec le concept planète. La notion de classe collective doit pallier cette lacune».

⁶ Pour plus de détails et des développements sur le système complet de types de la GAC, c.f. (Desclés 1990).

⁷ L'application se fait alors en deux étapes. Par exemple la localisation de Jean à Paris se formalise de la façon suivante. Nous avons les types suivants : J : Jean; L : Paris et FJFLH pour cette localisation.

La localisation s'applique d'abord à Jean pour retourner un opérateur de type FLH :

(FJFLH : localisation J : Jean) [localisation_Jean] : FLH

Le résultat est un opérateur de type FLH qui s'applique à un lieu Paris pour retourner une valeur de vérité v de type H :

([localisation_Jean] : FLH L : Paris) H.

3. VERS UNE TYPOLOGIE DES RELATIONS

Dans la GAC, les relations sont classées dans deux grandes catégories disjointes : les relations statiques et les relations cinématiques/dynamiques. Dans le cadre de cet article, nous n'aborderons que les relations statiques, car ce sont celles que nous testons actuellement dans les trois projets d'aide à la construction de terminologies que nous avons mentionnés.

Les relations statiques construisent des situations statiques. Les situations statiques restent stables pendant un intervalle temporel où ni le début, ni la fin ne sont envisagées. Par exemple, l'énoncé *Paris est en France*, correspond à une situation statique dans laquelle est décrite la position d'une entité (individuelle) par rapport à une autre (un lieu). Parmi les relations statiques, nous utilisons les relations de repérages (d'une entité par rapport à une autre) pour construire des terminologies.

3.1 Le repérage

Le repérage est un schéma général de relation : une entité X — entité repérée — est repérée par rapport à Y — une entité repère. Le repérage se spécifie suivant les propriétés algébriques qui lui sont attribuées axiomatiquement en diverses relations : identifications, différenciations, ruptions (figure 2).

La relation **d'identification** (qui se paraphrase en «X est identifié à Y», c'est-à-dire : l'entité Y sert d'identificateur pour l'entité X) est une relation binaire symétrisable et réflexive⁸. Elle est exprimée dans des énoncés tels que : *Paris est la capitale de la France*.

La relation de repérage qui est engendrée par **différenciation** (qui signifie «X est différent de Y») est non symétrique. C'est une relation très générale que nous détaillons par la suite.

La **ruption**, qui se lit «X est en rupture avec Y» (il n'y a aucune propriété commune entre X et Y) est une relation de repérage non réflexive et symétrique. Elle est exprimée dans des énoncés comme : *Il y a incompatibilité entre protons et électrons*. Très souvent, cette relation opère sur des classes disjointes issues d'une même classe par des attributions : *nombres pairs/ impairs, valeurs positives/négatives, quantitatif/qualitatif, etc.*

Les relations sémantiques s'insèrent ainsi dans un système de significations de repérage entre entités (Desclés 1982, 1987; Jouis 1993).

⁸ L'identification se spécifie en plusieurs relations telles que l'égalité extensionnelle et l'égalité intensionnelle. Sur ce point, voir (Desclés 1987).

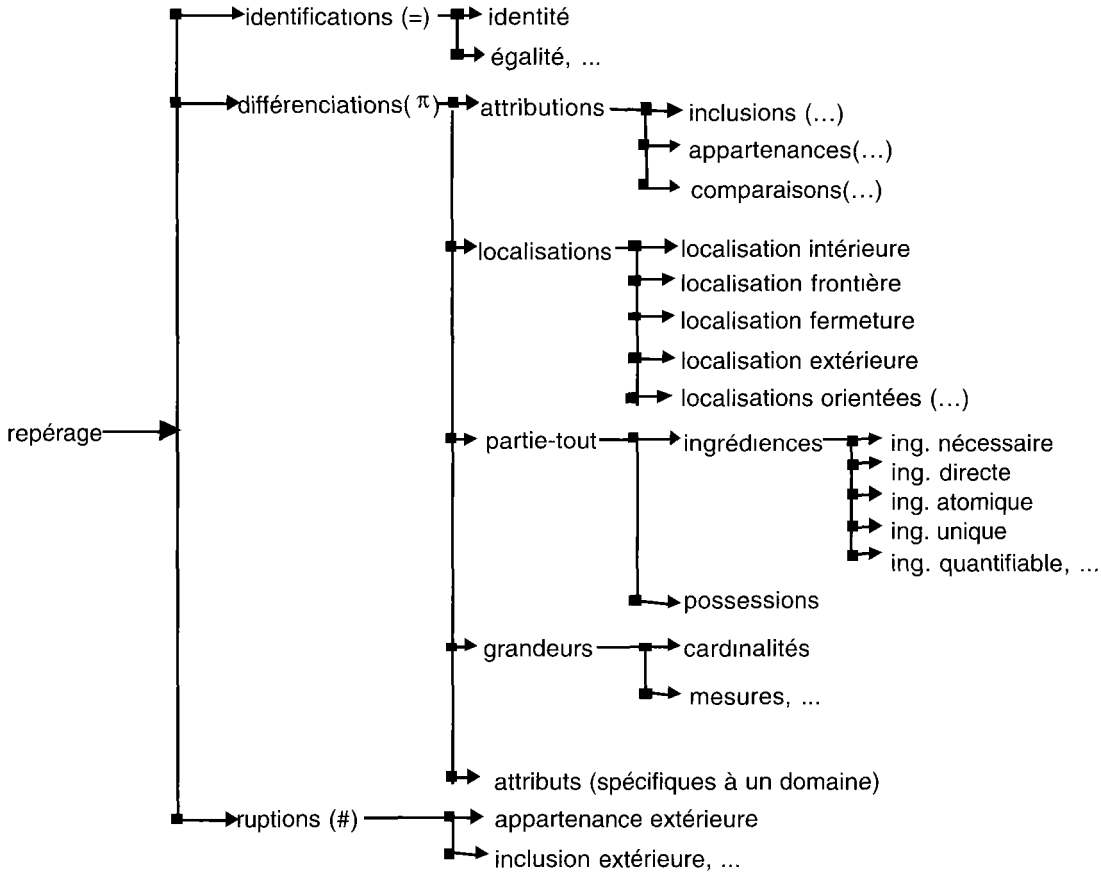


Figure 2 : Schéma de spécification des relations statiques

Dans ce système, la sémantique de chaque relation correspond ainsi à des propriétés intrinsèques⁹ :

- (1) son type fonctionnel (type sémantique des arguments de la relation);
- (2) ses propriétés algébriques (réflexivité, symétrie, transitivité, etc.);
- (3) ses propriétés de combinaison avec les autres relations dans un même contexte (contexte définitoire par exemple);
- (4) la relation plus générale et les relations plus spécifiques (une relation peut se spécifier en d'autres relations plus précises en fonction de propriétés nouvelles par rapport à la relation plus générale).

⁹ Ce qui signifie qu'elles sont indépendantes d'un domaine de connaissance particulier.

3.2 Les relations de différenciation

Parmi les relations issues de la différenciation, nous avons l'**attribution**, caractérisée par la non-symétrie. Elle se spécifie en :

- l'**appartenance** d'une entité individualisable (J) à une classe distributive (D)¹⁰, de type FJFDH, qui est non réflexive, non symétrique et non transitive¹¹. Elle est exprimée dans des énoncés tels que : *PI est un nombre réel*;
- l'**inclusion** entre classes distributives (D), de type FDFDH, qui est réflexive, antisymétrique et transitive¹².

La **comparaison** correspond à une relation d'ordre stricte (c'est-à-dire qu'elle est irréflexive, asymétrique et transitive) portant sur les entités individuelles : son type est donc FJFJH. Elle se spécifie en deux relations : supérieur (>) et inférieur (<).

La relation **partie-tout** est une relation générale permettant la décomposition d'un objet en ses éléments constitutifs. À travers cette relation, toute entité individuelle est considérée comme une unité complexe organisée. La relation partie-tout admet deux arguments qui sont respectivement l'objet décomposé et l'objet composant. Son type est donc FCFxH où x est de type J ou de type C. Les relations partie-tout comprennent les relations d'ingrédience et de possession.

La relation d'**ingrédience** est transitive et réflexive mais non symétrique, ce qui la différencie de l'inclusion. Elle est exprimée dans des énoncés comme : *La main fait partie du bras; Le fluor entre dans la composition des os des dents*. L'ingrédience se spécifie en plusieurs relations. En effet, il existe un grand nombre de propriétés décrivant les rapports entre l'objet composant et l'objet global¹³. Citons par exemple :

- l'**ingrédience nécessaire/non nécessaire** (*Le processeur est un des composants indispensables d'un ordinateur* vs *Le lecteur CD-ROM est un composant accessoire d'un ordinateur*); les caractéristiques nécessaire et non nécessaire acceptent la transitivité de la relation d'ingrédience;

¹⁰ Ou l'attribution d'une propriété une entité individuelle. Par exemple, l'énoncé *Socrate est un homme* signifie que l'entité individuelle *Socrate* appartient à la classe des *hommes* ou encore que le concept *être-un-homme* s'applique à *Socrate*.

¹¹ L'identification se spécifie en plusieurs relations telles que l'égalité extensionnelle et l'égalité intensionnelle. Sur ce point, voir (Desclés 1987).

¹² Comme dans l'énoncé *les hommes sont mortels*. Il est à noter que dans beaucoup de thésauri ou de modèles de réseaux sémantiques, on n'utilise que la relation générale «is-a» («est-un») sans distinguer l'appartenance de l'inclusion. Or, il y a une différence fondamentale puisque la première n'est pas transitive tandis que la deuxième est transitive, ce qui permet l'héritage de propriétés !

¹³ Voir Agata Jackiewicz-Desbertrand : *Contribution aux problèmes de l'extraction des connaissances : manifestations linguistiques et représentation informatique de la relation d'ingrédience*, Mémoire de DEA Sciences Cognitives, EHESS, 1992.

L'ingrédience directe/non directe (*L'opium figure parmi les ingrédients primaires de la Lamaline vs une molécule est constituée de neutrons, de protons et d'électrons, qui se regroupent en atomes*); un objet partie OP est un ingrédient direct de l'objet tout OT, s'il n'existe pas d'objet OP1 (différent de OP) tel que l'objet OP est ingrédient de l'objet OP1 et l'objet OP1 est ingrédient de l'objet OT. Sinon OP est un ingrédient non direct (fig. 3). L'ingrédience non directe est transitive, tandis que l'ingrédience directe est non transitive;

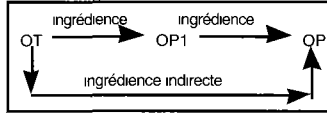


Figure 3 : Ingrédience directe / indirecte

- **L'ingrédience atomique/non atomique** (*Le plus petit élément constitutif d'un programme est le bit vs un ouvrage se décompose en chapitres, qui eux-même se décomposent en paragraphes...*); la caractéristique atomique de la relation d'ingrédience n'admet pas la transitivité, alors que la caractéristique non atomique l'autorise;
- **L'ingrédience unique/non unique** (*Une étoile jeune est constituée exclusivement d'atomes d'hydrogène vs L'air pur est un mélange de plusieurs gaz, dont les principaux sont l'oxygène et l'azote*);
- **L'ingrédience quantifiable/non quantifiable** (*La main est composée de cinq doigts; Chaque cellule humaine contient 46 chromosomes vs L'eau est constituée d'atomes d'oxygène et d'atomes d'hydrogène*); etc.

Les relations de **localisation** par rapport à un lieu sont exprimées par exemple dans les énoncés suivants : *Paris est en France, Un jardin entoure la maison, Le livre est sur la table...* Un X (un localisé) est localisé par rapport à Y (le localisateur). Les relations de localisation sont de type : FxFLH où x est de type J ou de type L, suivant le contexte du localisé¹⁴.

Des primitives de position peuvent être définies en faisant appel à quelques concepts rudimentaires de la topologie générale. Un lieu est alors visualisé soit dans son intériorité, soit dans son extériorité (excluant son intériorité et ses frontières), soit dans sa globalité (les limites et l'intériorité). On introduit les opérateurs de détermination topologique d'un lieu x in(x), ex(x), fr(x) et fe(x), déterminant respectivement l'intérieur, l'extérieur, la

¹⁴ Chaque occurrence d'un objet, dans un environnement pragmatique particulier, détermine un lieu (ou voisinage au sens topologique).

frontière et la fermeture de x . Les propriétés de ces quatre opérateurs¹⁵ nous permettent de spécifier les quatre relations binaires de localisation suivantes¹⁶ :

- loc-in : «être-dans»,
- loc-ex : «être-hors-de»,
- loc-fr : «être-à-la-frontière-de»,
- loc-fe : «être-à-la-fermeture-de».

Citons quelques propriétés de ces relations. La relation **loc-in** (exemple : «*Le boîtier contient la carte électronique*») est transitive, antisymétrique et non réflexive. La relation **loc-ex** (exemple «*La cuve de refroidissement se trouve à l'extérieur de la zone radioactive*») est irreflexive. La relation **loc-fr** (exemple : «*La cellule est délimitée par sa membrane*») est incompatible avec l'extérieur et l'intérieur et plus précise que la fermeture. La relation **loc-fe** (exemple : «*Jean est à Paris*») est incompatible avec l'extérieur et redondante avec la frontière et l'intérieur.

La relation **attribut** permet d'isoler des relations statiques qui échappent à une définition générale, c'est-à-dire qui sont liées à un domaine particulier. Par exemple, la relation «est-père-de», que l'on utiliserait en généalogie, ne peut pas être considérée comme une relation générale. Autrement dit, l'ensemble structuré des relations que nous proposons est considéré comme une ensemble d'invariants sémantiques, indépendant d'un domaine considéré, mais non exhaustif. Notre approche n'exclue pas la nécessité de rajouter des relations spécifiques au domaine.

4. CONCLUSIONS

La typologie des relations que nous proposons s'appuie fortement sur le modèle linguistique de la GAC, et plus particulièrement sur un sous-ensemble de primitives sémantiques qui se voudraient être indépendant d'un domaine de connaissance. Ces primitives sont organisées à partir de types sémantico-logiques dans un système de significations. Ce système est construit progressivement, en partant d'une relation très générale entre unités, pour arriver à des relations sémantiques de plus en plus précises en spécifiant au fur et à mesure des propriétés attribuées :

- type fonctionnel (type sémantique des arguments);
- propriétés algébriques (réflexivité, symétrie, transitivité, etc.);
- propriétés de combinaison avec les autres relations dans un même contexte;
- insertion dans un réseau de relations.

Dans ce système, la sémantique de chaque relation correspond ainsi à des propriétés intrinsèques, indépendantes du domaine.

¹⁵ Pour tout lieu x , nous avons par exemple :

$$\begin{aligned} \text{in}(x) & \dot{\bar{I}} x \dot{\bar{I}} \text{fe}(x). \\ \text{fr}(x) & \dot{\bar{I}} \text{fe}(x) \text{ (en effet } \text{fe}(x) = x \dot{\bar{E}} \text{fr}(x)) \\ x & \dot{\bar{C}} \text{ex}(x) = \bar{A} \\ \text{fr}(x) & = \text{co}(\text{in}(x)) \dot{\bar{C}} \text{co}(\text{ex}(x)) \end{aligned}$$

¹⁶ Sans rentrer dans les repères quantitatifs, nous pouvons distinguer de la même façon des localisations orientées en introduisant les primitives gauche(x), droite(x), devant(x), derrière(x), haut(x), bas(x). Toutefois, ces primitives ne peuvent être définies que si l'objet repère possède une orientation intrinsèque : *le devant de la maison, la droite d'un bateau (babord)*, etc.

Toutefois, si l'ensemble structuré des relations que nous proposons est considéré comme un ensemble d'invariants sémantiques, indépendant d'un domaine considéré, nous ne prétendons pas qu'il soit complet. Notre approche n'exclut pas la nécessité de rajouter d'autres relations plus précises ou spécifiques au domaine.

D'autre part, nous n'avons pas abordé dans cet article la représentation des entités atypiques. Enfin, nous n'avons décrit que les aspects statiques. Il est nécessaire d'y ajouter des **relations dynamiques** entre unités terminologiques. Ces relations permettent de décrire des processus ou des événements d'un domaine de savoir (unités terminologiques caractérisant des situations cinématiques ou dynamiques). Citons, par exemple, la représentation des mouvements spatio-temporels (Abraham 1995; Le Priol, en cours).

RÉFÉRENCES

- ABRAHAM, M. (1995) : *Analyse sémantico-cognitive des verbes de mouvement et d'activité: Contributions méthodologique à la constitution d'un dictionnaire informatique des verbes*, Thèse de doctorat EHESS, Paris.
- BISKRI, I. (1995) : *La Grammaire Catégorielle Combinatoire Applicative dans le cadre de la Grammaire Applicative et Cognitive*, Thèse de Doctorat, EHESS, Paris.
- BEAN, C. (1996) : «Analysis of Non-Hierarchical Associative Relationships Among Medical Subject Headings (MeSH): Anatomical and Related Terminology», *Knowledge Organization and Change*, ISKO 4th Conference, Washington DC.
- BÉGUIN, A., JOUIS, C. et W. MUSTAFA (1997) : «Évaluation d'outils d'aide à la construction de terminologies et de relations sémantiques entre termes à partir de corpus», *Actes des Premières Journées Scientifiques et Techniques (JST'97)*, Avignon, FRANCIL, AUPELF-UREF, avril 1997, pp. 419-426.
- DESCLÉS, J.-P. (1990) : *Langages applicatifs, langues naturelles et cognition*, Paris, Hermès.
- DESCLÉS, J.-P. (1987) : «Réseaux sémantiques : La nature logique et linguistique des relateurs», *Langages*, n° 87, pp. 55-78.
- DESCLÉS, J.-P. et C. FROIDEVAUX (1982) : «Axiomatisation de la notion de repérage abstrait», *Mathématiques et Sciences Humaines*, 78, pp. 73-119.
- FELBER, H. (1987) : *Manuel de terminologie*, Paris, UNESCO.
- GREEN, R. (1996) : «Development of a Relational Thesaurus in Knowledge Organization and Change», ISKO 4th Conference Proceedings, Washington DC July 15-19, *Advances in Knowledge Organization*, Vo.5 (AKO-5), Frankfurt, Index Verlag, pp. 72-79.
- GRIZE, J.-B. (1973) : *Logique moderne*, Fascicule II. Paris, Mouton/Gauthier-Villars.
- HUMBLEY, J. (1997) : site Web : <http://www-lli.univ-paris13.fr/>.
- ISO 704, (1987) : *Principes et méthodes de la terminologie*, Genève, Organisation internationale de normalisation (ISO/TC 37).

- ISO 1087 (1990) : *Terminology - Vocabulary = Terminologie - Vocabulaire*, Genève, Organisation internationale de normalisation.
- JOUIS, C. (1993) : *Contributions à la conceptualisation et à la modélisation des connaissances à partir d'une analyse linguistique de textes. Réalisation d'un prototype : le système SEEK*, Thèse de Doctorat. Ecole des Hautes Études en Sciences Sociales, Paris.
- JOUIS, C., BISKRI, I., DESCLÉS, J.P., LE PRIOL, F., MEUNIER, J.-G., MUSTAFA, W. et G. NAULT, (1997) : «Vers l'intégration d'une approche sémantique linguistique et d'une approche numérique pour un outil d'aide à la construction de bases terminologiques», *Actes des Premières Journées Scientifiques et Techniques (JST'97)*, Avignon, FRANCIL, AUPELF-UREF, avril 1997, pp. 427-432.
- JOUIS, C. et W. MUSTAFA-ELHADI (1996) : «Vers un nouvel outil interactif d'aide à la conception de dictionnaires électroniques spécialisés», *Lexicomatique et dictionnairiques. Actes des IV^{es} Journée scientifique du réseau thématique «Lexicologie Terminologie, Traduction»*, Lyon, sept. 95, Beyrouth. AUPELF-UREF et FMA, pp. 255- 266.
- LE PRIOL, F. (en cours) : *Extraction des relations dynamiques à partir de textes par exploration contextuelle*, Thèse de Doctorat en cours, Université Paris-Sorbonne
- LERAT, P. (1989) : «Les fondements théoriques de la terminologie», *La banque des mots*, numéro spécial, pp. 51-62.
- LERAT, P. (1990) : «L'hyponymie dans la structuration des terminologies», *Langages*, n° 98, pp. 79-86.
- LERAT, P. (1995) : *Les langues spécialisées*, Paris, PUF, coll. «Linguistique nouvelle».
- MEYER, I. et B. MCHAFFIE (1994) : «De la focalisation à l'amplication : nouvelles perspectives de représentation des données terminologiques», Clas, A. et P. Bouillon (dir), *TA-TAO : Recherches de pointe et applications immédiates, Troisièmes Journées scientifiques du réseau thématique de recherche «Lexicologie, terminologie et traduction»*, Montréal, 30 septembre, 1^{er} et 2 octobre 1993, Beyrouth, FMA et AUPELF-UREF, pp. 425-440.
- MIEVILLE, D. (1984) : *Un développement des systèmes logiques de Stanislaw Lesniewski*. Berne-Francfort s/ Main- New York, Peter Lang.
- MILLER, G. A. (1990) : «Nouns in Wordnet: a Lexical Inheritance System», *International Journal of Lexicography*, vol. 3, n° 4, pp. 245-264.
- MOLHOLT, P. (1996) : «Standardization of Interconcept Links and their Usage in Advances in Knowledge Organization and Change», ISKO 4th Conference Proceedings, Washington DC July 15-19, *Advances in Knowledge Organization, Vo.5 (AKO-5)*, Frankfurt/Main, INDEX Verlag, pp. 65-71.
- REY, A. (1992) : *La terminologie. Noms et notions*, 2^e édit. corrigée, Paris, PUF, coll. «Que sais-je ?», n° 1780.

- SAGER, J. (1990) : *A practical course in Terminology Processing*, Amsterdam/Pliladelphia, John Benjamins.
- VAN CAMPENHOUDT, M. (1996) : *Abrégé de terminologie multilingue*, accessible à partir du Site Web <http://www.refer.fr/termisti>.
- VAN CAMPENHOUDT, M. (1994) : «Les relations notionnelles expérimentées dans les microglossaires de TERMISTI : du foisonnement à la régularité», Clas, A. et P. Bouillon (dir), TA-TAO : recherches de pointe et applications immédiates. *Troisièmes Journées scientifiques du réseau thématique de recherche «Lexicologie, terminologie et traduction»*, Montréal, 30 septembre, 1^{er} et 2 octobre 1993, Beyrouth, FMA et AUPELF-UREF, pp. 409-423.
- WIELINGA, B., SCHREIBER, A. & J. BREUKER (1992) : «KADS: a modeling approach to knowledge engineering», *Knowledge Acquisition*, 4 (1), pp. 5-53.
- WINSTON, M.E., CHAFFIN, R. & D. HERRMANN (1987) : «A Taxonomy of Part-Whole Relations», *Cognitive Science*, vol. 11, n° 4, pp. 417-444.
- WÜSTER, E. (1981) : «L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et la science des choses», Rondeau, G. et H. Felber (dir), *Textes choisis de terminologie. Vol. I : Fondements théoriques de la terminologie*, Québec, Université Laval et GIRSTERM, pp. 55-113.

LE SIGNIFIÉ ET LE CONCEPT DANS LA DÉNOMINATION

Marcel DIKI-KIDIRI

CNRS (LLACAN), France

INTRODUCTION

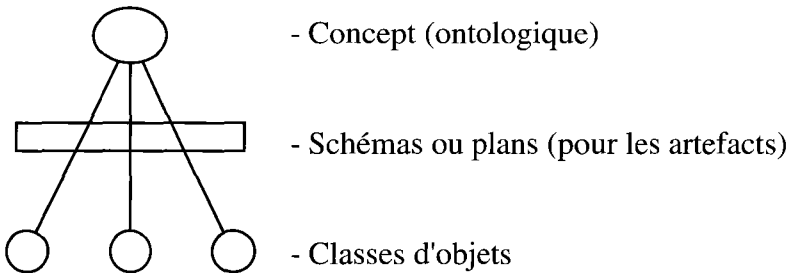
En travaillant sur l'instrumentalisation terminologique des langues africaines pour l'expression de réalités modernes qui n'existaient pas ou n'étaient pas prééminentes dans le monde des cultures traditionnelles africaines, nous sommes quotidiennement confrontés à la complexité de l'unité terminologique, à savoir le terme. Un certain nombre de contraintes s'imposait d'emblée à nous qui cherchions à dénommer des réalités nouvelles dans nos langues :

- a) Nous sommes ici dans un cas de figure où les réalités nouvelles à dénommer existent indépendamment de notre culture, étant le plus souvent créées en Occident. Ici, nous nous sommes interrogés d'une part sur la nature de ces réalités nouvelles, et d'autre part sur comment elles ont été dénommées dans les cultures occidentales où elles ont été créées. Cette démarche nous a conduit à mieux comprendre comment nous pourrions les dénommer dans nos langues respectives.
- b) Un deuxième cas de figure est celui de réalités anciennes en Afrique, mais qui n'avaient pas jusque là fait l'objet d'études avancées dans les cultures africaines traditionnelles. Au contact de l'Occident, la connaissance de certaines de ces réalités devient importante pour le développement économique, par exemple, les insectes ravageurs de coton (Henry Tourneux 1993). Ici, nous procédons à trois types d'observation : la nature intrinsèque de ces réalités, les dénominations occidentales, et les dénominations traditionnelles africaines. L'évaluation de celles-ci permet d'orienter l'activité de mise à niveau terminologique.
- c) Dans tous les cas de figure, les termes produits, qu'ils soient néologiques ou non, sont des unités linguistiques censées faire partie intégrante d'une langue naturelle, et destinées à servir à la communication aussi bien dans le cadre d'activités professionnelles (ex. culture du coton) que dans celui du transfert des connaissances (ex. scolarisation, alphabétisation, formation). *Le terme est donc un signe linguistique qui permet, dans une langue naturelle, l'expression et la communication d'une connaissance spécialisée.*

Si donc le terme est bien un «signe linguistique» à la saussurienne, force est de reconnaître que la seule relation binaire signifiant/signifié est insuffisante pour décrire le terme dans toute sa complexité. Pour en rendre compte de manière plus précise, il importe de distinguer les notions de concept et de signifié, de sorte que le «terme» puisse être décrit dans plusieurs dimensions concentriques, ainsi que le suggère Teresa Cabré¹, qui distingue la dimension linguistique, la dimension sociale et la dimension utilitaire. En ne restant que sur la dimension linguistique, nous pensons qu'il y a encore lieu de distinguer trois facettes ou axes qui s'articulent autour du signifiant, du signifié et du concept. Si toutes les questions relatives à la forme (ou aux différentes formes) du terme se traitent au niveau du signifiant, c'est au niveau du signifié que se situent la problématique de la construction du sens, de la reconceptualisation en termes de perceptions culturelles, de polysémie. Enfin, c'est au niveau du concept que l'on se pose l'essentiel des questions relatives à la cognition et à la nature du savoir, de son objectivité et de son universalité vraies ou vraisemblables. Lorsque l'on est amené à faire une distinction entre les mots ordinaires de la langue dite générale et les termes propres à une spécialité, un domaine avancé du savoir, il y a lieu de prendre ensemble cette triple articulation du terme autour du signifiant, du signifié et du concept.

CONCEPT ET CLASSE D'OBJETS

Si nous considérons un objet fabriqué, par exemple une maison, un ordinateur, une bicyclette, un pot, etc., on dira que cet objet a été conçu avant d'être réalisé. Cette conception peut se traduire en une série de réflexions, d'études, de calculs, etc., aboutissant à une représentation mentale idéale de ce qu'*est* l'objet. C'est cette représentation mentale que nous appellerons ici le «concept» de l'objet. Ce concept peut être réduit sous une forme écrite qu'on appelle «plan de réalisation». Chaque objet réalisé grâce à ce plan est une instance concrétisée du concept. Outre que tous ces objets partagent les attributs essentiels du concept, chacun d'eux peut avoir des attributs propres dits «personnalisés» ou «localisés», qui peuvent justifier une structuration en sous-classes d'objets. Cette variation commence d'ailleurs souvent dès le niveau de la conceptualisation en donnant une famille de concepts. Pensez à la production d'une gamme de voitures qui prévoit dès la conception plusieurs modèles dans la même gamme et plusieurs variantes dans le même modèle.



Parce que ce type de concept est à la base de la création (production, construction, existence) des objets ainsi conçus, le décrire revient à décrire ce que sont ces objets en

¹ Voir bibliographie à la fin du texte.

eux-mêmes, c'est-à-dire leur ontologie. C'est pourquoi, si nous devons caractériser fondamentalement ce type de concept, nous préférons l'expression *concept ontologique* aux autres terminologies florissantes de la non moins abondante littérature scientifique sur le sujet.

L'expression *concept ontologique* convient également à cet autre cas de figure où il s'agit de caractériser des objets naturels non créés par l'homme : animaux, végétaux, minéraux, esprits, astres, «forces de la nature» etc. Ici la question centrale est la catégorisation, comme le montre très bien toutes les recherches sur la cognition, la prototypicalité etc. Typiquement, lorsqu'on trouve un animal inhabituel, le degré de connaissance encyclopédique que l'on peut rassembler sur lui permet de catégoriser avec d'autant plus de précision que ce degré est élevé. Ainsi, la baleine avait d'abord été classé comme un poisson avant de rejoindre la catégorie des mammifères. Puisque l'homme n'a pas créé la baleine, il n'en reconstituera qu'imparfaitement le concept ontologique à la hauteur de la somme de connaissances qu'il peut avoir de lui, en extrapolant vers l'idéal à partir de ce qu'il sait. C'est ce que Lakoff (1987 : 68) qualifie si justement de *modèle cognitif idéalisé*.

Qu'on chemine du concept vers les objets (artefacts) ou des objets (espèces naturelles) vers le concept, le cognitiviste et/ou l'ontologiste se préoccupe d'expliquer aussi complètement que possible, la réalité en soi. C'est la dichotomie *concept/référent* qui est ici le thème de la recherche, l'objet de l'étude. Si cette étude intéresse le terminologue, c'est en autant qu'elle lui permet de décrire, par exemple dans une définition en terminographie, les composantes essentielles (ou prototypiques) d'un objet de connaissance spécialisée, qu'il s'agisse d'un artefact ou d'une espèce naturelle. Il s'agit ici d'une problématique de caractérisation et de catégorisation. Mais la question de la *dénomination* proprement dite sera traitée ailleurs, dans le cadre du signe linguistique ou terminologique et en relation avec une dimension perceptuelle.

DÉNOMINATION ET PERCEPTIONS CULTURELLES

Il est courant qu'un produit porte un nom de code au cours de sa conception et un autre nom, «commercial» cette fois, lors de sa mise sur le marché, sans compter celui que les utilisateurs (consommateurs) lui donneront. L'affaire devient encore plus complexe lorsque la terminologie est envisagée sous un angle multilingue ! Nous commencerons donc par présenter des cas simples afin de bien mettre en lumière les phénomènes que nous voulons décrire.

Polysémie et organisation prototypique du sens

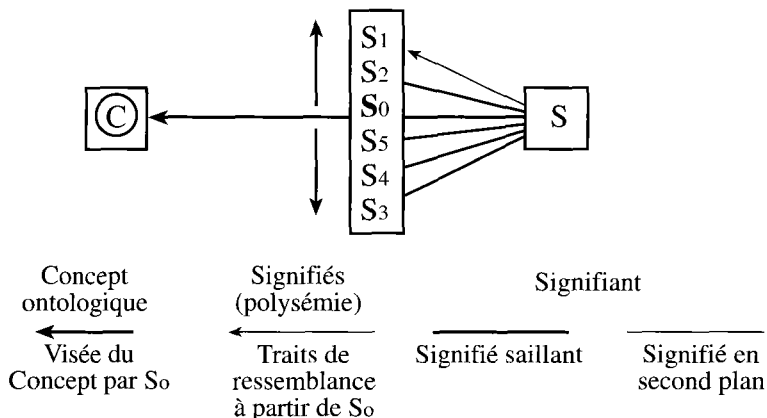
Le mot *souris* en français désigne prioritairement (ou prototypiquement) un «quadrupède de la famille des rongeurs, appartenant au genre rat» (Littré 1994). Mais on lit dans le même article plusieurs autres sens figurés, dérivés, métaphoriques (etc.) tels que :

- un homme qui a très peur ou qui éprouve un grand embarras;
- une couleur gris-argenté;

- un muscle charnu qui tient par un bout à la manche du gigot;
- l'espace qui est dans la main entre le pouce et l'index, etc.

Il est à noter que ces deux derniers exemples relèvent du même domaine, l'anatomie. En informatique, comme on le sait, le mot désigne un *dispositif électromécanique ou électro-optique de pointage et de saisie*. C'est sur cette polysémie qui va insensiblement de la langue générale à la langue de spécialité que nous allons nous concentrer. Il est facile d'expliquer pourquoi ces différents sens du mot «souris», notamment par l'analogie, la métonymie, la métaphore, etc. (cf. Lakoff et Johnson 1985)². Par ailleurs, la polysémie en tant que phénomène linguistique est suffisamment bien étudiée pour qu'on s'y attarde ici pour elle-même. Nous nous contenterons donc de souligner quelques points pertinents pour la dénomination.

Du point de vue de la catégorisation, axe concept/référent, seul le petit animal désigné par le mot «souris» peut être classé comme un «rongeur». Les autres signifiés ne le peuvent pas. Un homme, même apeuré, n'est pas un rongeur, pas plus qu'un muscle mobile, etc. Par contre, l'homme apeuré ou embarrassé a un comportement qui, dans la culture française (ou francophone), est perçu comme similaire au comportement du rongeur. L'espace réduit entre le pouce et l'index a été perçu, toujours dans la même culture, comme évoquant un trou de souris et a été donc appelé ainsi. La couleur gris-argenté étant dominant chez les souris, la synecdoque se comprend. Pour le muscle comme pour le dispositif électronique, c'est la ressemblance de forme qui a suffi à permettre le rapprochement. Tous les signifiés dérivés n'ont aucun trait commun entre eux, mais partagent, chacun, au moins un trait commun, plus ou moins symbolique, avec le signifié basique (prototypique) qui renvoie au concept ontologique de «quadrupède rongeur du genre rat» désigné par le signifiant *souris*.



La position basique (prototypique) du signifié (S₀) réflexe du concept ontologique (C₀) «quadrupède rongeur...» que désigne le signifiant (S) *souris* peut varier avec le temps. Pour une génération d'enfants vivant dans une grande ville occidentale comme

² Voir bibliographie en fin de texte.

Paris, et qui ont grandi avec les jeux vidéo, le mot *souris* évoque prioritairement (donc basiquement, prototypiquement) l'accessoire informatique. Et cette acception (signifié S₅) supplante et relègue au second plan l'acception première d'animal «quadrupède rongeur».

Concepts et signifiés ou percepts

La distinction entre *concept* entendu comme *ontologique* et *signifié* envisagé comme le lieu des perceptions culturelles s'impose encore plus nettement lorsqu'on compare la dénomination d'un même *artefact* dans plusieurs langues.

La bicyclette est appelé

- *gbâzâbângâ* «roues de caoutchouc» en sängö (Centrafrique);

- *nàgàsó*³ «cheval de fer» en bambara (Mali);

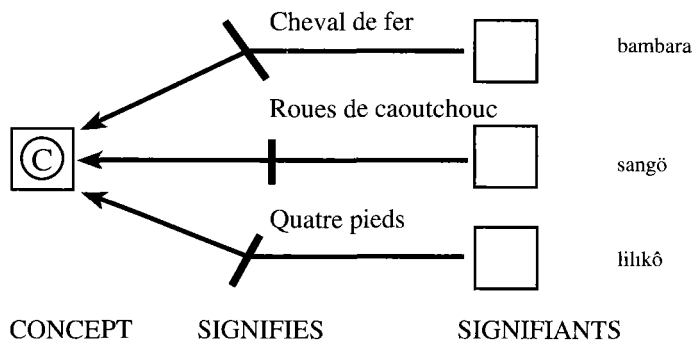
- *magu-mâkwanganya*⁴ «quatre pieds» en lilikô (langue bantou de la zone D au Congo Démocratique).

Ces différentes appellations témoignent à la fois de la diversité dans la perception de l'objet bicyclette par des communautés de langues et de cultures différentes, ayant des passés différents. Les Centrafricains connaissaient la roue et avaient été, par ailleurs, soumis aux travaux forcés de la récolte du caoutchouc végétal dès les premières années de la colonisation de leur pays (vers 1910). Les roues de caoutchouc (pneu) de la bicyclette ont donc retenu particulièrement leur attention et motivé leur choix dans la dénomination de ce véhicule. De leur côté, les Bambara, qui connaissaient le cheval, ont perçu une ressemblance entre cet animal et la bicyclette : les deux se montent à califourchon et permettent à l'homme de se déplacer plus rapidement qu'à pied. Par contre, seul ce dernier trait (déplacement plus rapide qu'à pied) a retenu l'attention des Bolikô du Congo Démocratique. Leur appellation «quatre pieds» s'entend comme «le véhicule qui dédouble vos pieds» et qui vous permet donc d'aller deux fois plus vite. Dans tous les cas, ces appellations ne cherchent pas à rendre compte de la structure schématique (concept ontologique) de la bicyclette. Les différents signifiés attachés aux différents signifiants de ces dénominations représentations mentales correspondent à des *points de vue* différents sur l'objet, points de vue motivés par et dépendant du passé culturel propre à chaque communauté. Le signifié n'est donc pas l'équivalent du concept, mais seulement un pointeur vers le concept, un point d'ancrage qui permet de saisir globalement le concept, en soi beaucoup trop riche. Nous préférons pour cette raison l'appeler «*percept*».

Si on peut aussi employer le terme de «concept» à propos des signifiés, il faudra alors parler de «concepts perceptuels, interactifs, relationnels, etc.» pour bien les distinguer des «concepts ontologiques» décrit plus haut (Voir section 1 : Concepts et classes d'objets). Il est certainement plus clair de parler ici de *percepts*, ou à la rigueur de *signifiés*, et de réserver le terme de *concept* à la notion de «concept ontologique» (voir figure ci-après).

³ Information confirmée par Gérard Galtier.

⁴ Information donnée par Edema Atibakwa.



CONCLUSION

La dénomination apparaît ainsi comme fortement liée à une perception culturelle inscrite essentiellement dans la relation signifiant/signifié ou plus précisément *signifiant/percept*. Le concept ontologique étant une abstraction idéalisée de ce que sont les objets est beaucoup trop riche pour être totalement circonscrit dans des mots. Aussi, de cette richesse du concept, l'homme retient-il quelques traits, saillants à ses yeux, et qui motiveront sa dénomination du concept. La dénomination la plus adéquate, la mieux acceptée, n'est pas forcément celle qui rend le plus précisément le concept, mais bien celle qui s'intègre le mieux à la langue et à la culture de la communauté des locuteurs. Pour aller plus avant dans le sujet de la motivation, on peut se reporter à la communication de notre collègue Edema Atibakwa présentée ici même. Pour aller plus loin également, il faudrait creuser davantage la paradoxale relation entre la langue de spécialité et la langue générale à la jonction desquelles se situe le terme en tant qu'unité terminologique. En effet, le terme, on vient de le voir, se doit d'appartenir totalement à la langue dite générale, car c'est la langue tout court, et aussi de participer à ce sous-ensemble linguistique chargé d'exprimer des concepts propres à un domaine spécialisé du savoir. Cette double articulation du terme mériterait d'être étudiée plus avant.

RÉFÉRENCES

- CABRÉ, Teresa (1997) : «Éléments pour une théorie de la terminologie», TIA97, *II^{es} Rencontres terminologies et intelligences artificielles*, Université Toulouse-le-Mirail, 24 avril 1997 (sous presse).
- KLEIBER, Georges (1990) : *La sémantique du prototype, catégories et sens lexical*, Paris, PUF, 199 p.
- DUBOIS, Danièle (dir) (1991) : *Sémantique et cognition, catégories, prototype, typicalité*, Paris, Éditions du CNRS.
- LAKOFF, G. et M. JOHNSON (1985) : *Les métaphores dans la vie quotidienne*, Paris, Éditions de Minuit.
- LITTRÉ, Émile (1994) : *Dictionnaire de la langue française*, Paris, Éditions de la Fontaine au Roi.

MÉMOIRE DES MOTS ET DISCOURS : LA MÉMOIRE ÉNONCIATIVE

Khadiyatoulah FALL et Daniel SIMEONI

Université du Québec à Chicoutimi, Chicoutimi et York University, Toronto, Canada

Quel beau livre ne composerait-on pas en racontant la vie et les aventures d'un mot ? Sans doute il a reçu diverses impressions des événements auxquels il a servi; selon les lieux, il a réveillé des idées différentes; mais n'est-il pas plus grand encore à considérer sous le triple aspect de l'âme, du corps et du mouvement ?

Balzac (*in* Louis Lambert, 1842)

1. INTRODUCTION. UN CADRE DE TRAVAIL

Notre travail s'inscrit dans la mouvance des théories dites de «l'énonciation», cadre que John Lyons a su redéfinir à l'intention des linguistes non francophones et singulièrement de ceux des pays anglophones, par une glose particulièrement heureuse. Lyons (1982), parle à propos de cette école française de linguistique dont la généalogie court de Bally à Guillaume, puis de Benveniste à Culioli, de «structuralismes phénoménologiques post-saussuriens».

Le champ ainsi balisé est celui d'une linguistique de la parole laissée «en creux» par Saussure. Un certain nombre de conséquences découlent des choix de Saussure, parmi lesquelles la mise de côté du coup de force épistémologique grâce auquel le linguiste genevois a institué le concept de «langue» et, plus utilement, l'approfondissement du champ laissé libre par la construction de l'objet «langue».

Il ne s'agit pas dans le cadre de ce colloque de relire et de réévaluer la linguistique structurale. Nous nous attacherons toutefois ici à un élément distinctif, rapidement évoqué par Saussure (1982 : 102-103) à propos du principe «du caractère linéaire du signifiant» :

«Le signifiant, étant de nature auditive, se déroule dans le temps seul et a les caractères qu'il emprunte au temps : a) *il représente une étendue*, et b) *cette étendue est mesurable dans une seule dimension* : c'est une ligne [...] Tout le mécanisme de la langue dépend [de ce principe].

[...] Dans certains cas cela n'apparaît pas avec évidence. Si par exemple j'accentue une syllabe, il semble que j'accumule sur le même point des éléments significatifs différents. Mais c'est une illusion; la syllabe et son accent ne

constituent qu'un acte phonatoire; il n'y a pas dualité à l'intérieur de cet acte, mais seulement des oppositions diverses avec ce qui est à côté.»

Nous posons pour notre part que l'ordre de la parole et du discours déploie une syntagmatique de l'accumulation, ou de répétition par espèce d'empilement (ce que Culioli (1984) a du reste évoqué sous la forme d'un phénomène d'«Höherung» - ou «point d'accumulation»). Si la langue est une mise en ordre et que, de même que ses composantes minimales que sont les phonèmes, ses séquences syntagmatiques se succèdent logiquement (i) au sein de la phrase, (ii) phrase après phrase, la parole en revanche procède par cumul. La langue est le lieu d'une permanence mythique qui illustre une structure contrastive hors-histoire issue de l'imagination théorique du linguiste. La parole met en jeu un enchaînement qui ne se satisfait pas de critères seulement distinctifs : elle impose une mémoire non tant séquentielle que pluridimensionnelle et pour ce faire, emprunte à des moyens plus esthétiques que logiques.

Les objets étudiés dans un tel cadre sont du coup très différents de ceux dont traitent les linguistiques structurales conventionnelles. Au premier rang de ces objets d'un troisième type figurent : les particules énonciatives, les formes auxiliaires au sens large (y compris aux niveaux propositionnel et transphratique), et surtout les multiples avatars d'une modalisation étendue omniprésente, fonction réhabilitée par rapport à l'information idéationnelle du «message» transmis.

Nos recherches actuelles (Fall, Simeon 1993, 1995) sont compatibles avec certaines perspectives fonctionnelles-cognitives non francophones : tels les travaux de Halliday, Lakoff et, sur un plan plus formel, ceux de Langacker. Cette liste n'est nullement limitative. S'agissant de structures d'usage qui sont par définition des structures acquises, il importe toutefois de garder à l'esprit que toute cognition ancrée dans un sujet énonciateur est, *ipso facto*, culturelle et sociale. Les chercheurs anglo-saxons que nous venons de citer sont loin d'avoir le même regard sur cette dimension.

La perspective ici adoptée peut aussi s'intégrer à une sociosémiotique du langage, car le principe même de la parole est extérieur à la langue. Le lien avec certains travaux sociologiques comme ceux de Bourdieu devient alors possible :

«Quand il s'agit du monde social, les mots font les choses, parce qu'ils font le consensus sur l'existence et le sens des choses.» (Bourdieu 1994 : 138)

Simplement il faudra :

«...chercher *hors des mots*, dans les mécanismes qui produisent et les mots et les gens qui les émettent et les reçoivent, le principe du pouvoir qu'une certaine manière d'user des mots permet de mobiliser». (Bourdieu 1989 : 63)

Sur ces bases, il nous a paru que nous pouvions distinguer deux grands types de structures mnémophoriques, chacun s'inscrivant dans un «temps» distinct. Le premier, le mieux étudié et que nous n'évoquons ici qu'en passant, s'ancrerait dans l'étymologie pour mieux s'en distinguer et déboucher sur une stylistique et une lexicologie des effets expressifs. Les travaux fondateurs demeurent ceux de Charles Bally (1951) et de Henri Frei (1929). Le deuxième temps nous paraît plus occulté dans la mesure où l'accent mis sur

L'opposition synchronie/diachronie a pu faire oublier que les structures de la mémoire incorporée dans les mots et les syntagmes de l'usage sont des constructions prises dans des réseaux de signification certes construits mais jamais figés car toujours en négociation. Ce deuxième temps qui nous intéresse ici n'est autre que celui de la mémoire *en train de se faire*, ce mouvement perpétuel des mots dont l'histoire peut se lire au présent comme le pressentait déjà Balzac.

2. LE CORPUS. PRÉSENTATION

Par principe, nous travaillons sur des corpus conversationnels où les thèmes débattus révèlent une grande instabilité notionnelle. Les participants que nous rassemblons à cet effet sont invités à cerner les frontières de domaines maximale-ment ouverts au cours du processus normal de référenciation. Notre stratégie de recherche consiste à faire actualiser des lexèmes «sensibles» dans des situations elles-mêmes potentiellement conflictuelles.

Le corpus servant ici de support réunissait comme intervenants dix-huit (18) adolescents dont quatre (4) Québécois de souche plus un (1) adopté (à l'âge de quelques mois) et trois (3) étudiants étrangers dans le cadre d'un échange. Dix (10) de ces jeunes (en sus de l'adopté) étaient nés à l'étranger. Quatre (4) étaient issus de familles mixtes, et huit (8) avaient des parents étrangers d'une même origine. L'encadrement était assuré par des adultes en position de pouvoir statutaire : les deux interviewers étaient des universitaires auxquels s'était joint un animateur de pastorale.

Le thème discuté était celui de l'«identité»; les adolescents étaient invités à dire comment ils se percevaient du point de vue identitaire, soit comment ils s'identifiaient personnellement, et comment selon eux, les autres les percevaient.

Le contexte de l'échange (une seule rencontre d'une durée de soixante-quinze (75) minutes) était le suivant : le Québec, un an après un référendum d'autodétermination dont le résultat avait été extrêmement serré, et à la suite duquel des propos avaient été tenus par des leaders politiques au plus haut niveau, rendant responsables les Québécois non francophones et immigrants, de l'échec de cette tentative d'indépendance constitutionnelle.

Le bilan partiel dont nous faisons état ici s'insère dans une recherche à plus long terme sur les conditions de production des énoncés en situation d'interaction. Cette recherche s'articule avec d'autres, à l'entrecroisement de la linguistique et de la traductologie descriptive d'une part et de l'autre, au point de convergence de la linguistique, de la littérature et des sciences sociales.

3. OBSERVATIONS

Écartons d'emblée une fausse problématique. La question de la «sincérité» ou de la «fidélité» des propos tenus — comment se fier à ce que disent nos informants quand on sait que l'on ne prend pas les précautions d'usage ? — ne nous paraît pas pertinente. Dans le cadre de nos travaux en effet, nous ne nous intéressons pas tant aux contenus de pensée des locuteurs participants, qu'à la façon dont ces derniers produisent des flux discursifs soit heurtés, soit au contraire plus coulés. Ce sont les accidents de la parole que nous étudions,

accidents que les locuteurs ne contrôlent pas. Une donnée importante, sur ce point de la forme et de la façon générale dont les locuteurs expriment leurs idées, est que ceux-ci savent être intransigeants : si d'aventure l'interlocuteur introduit la moindre distorsion, voire un infléchissement, dans les formes utilisées par le locuteur, celui-ci n'hésitera jamais à redresser l'énoncé «fautif», et ce de façon parfois plus déterminée encore que si le fond du propos était sciemment altéré :

Question : eh tu as dit eh eh au Vietnam je me sens plus comme chez nous c'est ça que tu as dit ?

K : ben pas plus là mais j'me sens comme chez moi

Le propos de K., tel qu'il l'avait énoncé dans une séquence antérieure, était effectivement :

là j'me sens comme eh ... t'sé ch ch't'allé au Vietnam pi eh ... c'tait comme si j'étais comme chez nous.

3.1 L'enchaînement des stimulus

De la notion ouverte ici : [IDENTITÉ] à ses actualisations nominales et verbales, cette formulation recouvre un certain nombre de problèmes théoriques épineux, par exemple en ce qui concerne les questions de précedence cognitive : qu'est-ce qui est premier, du nom ou du verbe ? Trancher (ou non) nécessiterait un protocole particulier que nous n'abordons pas ici. Constatons simplement que d'une sollicitation relativement abstraite, l'entretien a glissé progressivement (de façon directive) vers des formulations plus dynamiques. La transition s'est faite de façon graduelle, la dénomination faisant place à des questions ouvertes et même fermées :

Qu'est-ce que c'est l'identité ?

L'identification

Le sentiment d'appartenance

Qui êtes-vous en dedans ?

Comment est-ce que vous vous sentez ?

Comment vous vous sentez en dedans ?

Comment vous vous identifiez ?

Comment vous vous considérez ?

Comment vous pensez que les autres vous reconnaissent ?

Comment les autres vous identifient aussitôt qu'ils vous voient ?

Comment est-ce que vos parents s'identifient ?

Est-ce que vous vous identifiez comme vos parents ?

On est passé ainsi graduellement d'une notion posée en essence («l'identité») à une dynamique d'identification plurielle libérant les interprétations possibles.

3.2 Confirmation de travaux antérieurs

3.2.1 Un certain nombre de jalons discursifs viennent baliser le déroulement du propos, comme en témoignent notamment certains marqueurs : «bon» ou «ben», «là» (dans sa double fonction, au Québec, de relance ou de bouclage), etc. Se trouvent ainsi

encadrés et mis en exergue de façon plus ou moins explicite certaines séquences thématiques et certains syntagmes focalisés. Le domaine une fois balisé, le débit tend à devenir plus fluide, mieux assuré, pleinement pris en charge.

A : Ben l'identité là c'est eh c'que chaque personne a ... pi que d'autres personnes ont pas c'est là ... pi c'est genre là te ... nos empreintes digitales là c'est notre identité à nous autres c'est propre à nous autres t'sé y a p ... y a personne qui peut n'avoir d'autres là ... c'est sûr qu'y a quelques ressemblances mais pas personne est pareil

[Réponse] : OK.

A : Moi j'dis qu'est ça notre identité

Le travail de rethématisation (voir infra 3.3.2.2) est bien jalonné, à chaque étape du processus de parole : un «ben» d'ouverture; quelques «là» de relance (trois premières occurrences glosables comme autant de «t'sé»); deux «là» de bouclage; une fréquence élevée de «c'est» et de «y a» visant à poser une nouvelle définition de la notion. Après stabilisation analogique (= l'identité c'est comme nos empreintes digitales), l'échange débouche sur un énoncé fluide, marqué de façon exemplaire :

Moi : terme de départ, thématisation du sujet énonciateur. Centrage sur EGO.

j'dis qu' : redite explicite du sujet et de sa fonction. LOQUOR.

c'est ça : thématisation emphatique de l'objet identifié par EGO.

notre : nouvelle appropriation, cette fois-ci collective (par désindividuation légitimante).

3.2.2 Comme nous l'avions déjà noté dans le compte rendu d'un précédent travail sur un corpus différent (Simeoni, Fall & Buyck 1993), le flux de la parole paraît plus heurté lorsque la représentation notionnelle perd de son homogénéité. Il peut en résulter des effets d'incohérence de surface tout à fait remarquables dont nul n'est à l'abri. L'exemple qui suit met en scène la déstructuration notionnelle de l'un des interviewers, associée à l'incapacité dans laquelle il se trouve de «comprendre», au sens propre du terme, ce que lui dit un informant :

- Question : ... par exemple bon je l'sais pas d'quelle eh ... de quelle religion est-tu ?

- J : j'sui baptisée ici

- Question : oui mais ta religion est-ce que tu es musulmane ou est-ce que tu

- J : non ... j'étais protestante mais j'me suis faite baptiser ici

- Interviewer : OK.

- J : là chu catholique

- Question : OK. tes parents sont protestants aussi ?

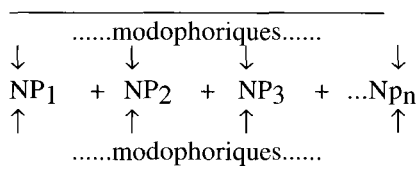
- J : ma mère non

À ce stade, la représentation de l'interviewer est si trouble qu'il s'ensuit une séquence remarquablement confuse :

- Question : OK. fait que là c'est ça t'sé les ... est-ce que dans la dans ton ton tes parents là ou peut-être eh ... étant donné que t'es protestante ici aussi là on est dans

un ... un milieu qui est eh chrétien ... mais eh ... il y a certaines valeurs des fois des des valeurs par exemple au niveau de la famille eh ... de nos grand-parents ...

3.2.3 Les participants se livrent en général à une modalisation effrénée, suivant en cela le schéma général d'une énonciation par empilement d'opérations imbriquées (Chambon & Simeoni 1997) :



L'introduction de ce schéma éclaire nos travaux antérieurs (Simeoni & Fall 1992; Fall, Simeoni, Chambon 1994) : l'enchaînement des noyaux propositionnels (ou lexis élémentaires) est indissociable de leur déploiement sous conditions de modalisation appropriées (intonation, modes, formes aspectuelles, jalons énonciatifs, structures auxiliantes, interjections, marques de validation, etc.). Or cette modalisation discursive ne s'exerce pas de façon linéaire. Elle impose aux participants sa prise en compte de façon diffuse et réitérée. Et cette réitération n'est ni redondante ni triviale; elle oriente la teneur du propos au même titre et parfois plus que la substance avérée du «message».

3.3 Nouvelles données révélées par la problématique de la «mémoire des mots»

Le deuxième temps mnémophorique : renvoie à une mémoire qui active des préconstruits notionnels, sans cesse reconstruits dans le temps de l'énonciation.

3.3.1 Articulation avec la problématique de l'anaphore

L'anaphore, signe manifeste d'une mémoire énonciative fondant le suivi du discours, ne sert jamais une fonction de réplication pure et simple. Le processus anaphorique consiste toujours en une rethématisation. De ce point de vue, on peut ranger l'anaphorisation parmi les opérations plus générales de redéfinition, ou de reconstruction du domaine notionnel. La reprise d'un terme, à l'aide ou non de formes délexicalisées, est ce qui fonde au bout du compte son acception sémantique (le consensus discursif).

3.3.2 Conditions de la production mémorielle en temps réel

3.3.2.1 Le point de départ de la reconstruction n'est jamais donné de façon absolue. Il semble qu'il y ait, de la part de l'énonciateur, le souci d'ordonner sa parole en tenant compte de l'état avéré ou supposé du domaine notionnel chez son ou ses vis-à-vis. Dans notre corpus, la reconstruction passe fréquemment par l'énonciation de ce que l'on nommerait, dans une perspective logiciste, *non-p*. Il y a en effet dans un premier temps *vidage* de ce qu'il faudrait peut-être considérer comme l'inscription du domaine dans un champ doxologique plus vaste, en tout cas un champ que l'énonciateur suppose préconstruit par la société ou les co-énonciateurs et, seulement après, *remplissage*, en fonction de ce que l'énonciateur cherche à obtenir, de ses convictions et de ses intérêts

propres, soit de son positionnement dans le champ doxologique (tel qu'il apparaît filtré par l'échange).

3.3.2.1.1 L'opération de *vidage* a plusieurs modalités : négation; concession suivie d'une rupture (ou décrochement) : «c'est sûr que... mais...»; atténuation ou euphémisation : «j'peux pas vraiment dire que...». Conformément à ce que pose notre schéma, ces modalités sont imbriquées dans la même proposition ou la même phrase.

(i) exemples de négation première

G. : (immédiatement après s'être présenté) ... pi eh j'veins pas de la région ... j'veins de [Y] ... mais eh ... j'me sens plus ... ap ben j'appartiens plus à ici [...] j-j'ai plusieurs tendances là ben ... j'dis pas qu'j'appartiens au Québec ... principalement nécessairement j'veux dire ... j'appartiens plus à aux communautés alentour de moi c'est pas eh

P. : pour ça t'sé y a pas vraiment d'conflit avec mes parents ... surtout genre peut-être avec la langue là t'sé ma mère elle aimerait ça que j'parle toujours eh t'sé bambara t'sé c'est ça qu'i parlent en Afrique ... ben ... pas tout en Afrique mais au Mali

L. : ben ... disons que c'est pas que je désire conserver mais c'est comme un peu ... un moule qu'on m'a donné

Il semble que chez certains locuteurs en situation polémique ou défensive, cette opération première occupe tout l'espace des possibles stratégiques. Une caricature à peine forcée est le tic de parole bien connu qui consiste à ouvrir systématiquement sur un refus, un rejet ou une mise à distance du propos d'autrui, quand bien même les représentations de l'interlocuteur ne se distinguent que minimalement des siennes propres.

(ii) exemples de concessives, préalablement à une prise en charge plus directe

Il s'agit de loin de la stratégie la plus courante, signalant un premier balisage *a contrario* du domaine avant rééquilibrage, rééquilibrage servant lui-même de préalable (éventuel) à une redéfinition positive :

R. : ben j'dirais pas qu'le Canada est tolérant moi en général ... bon peut-être que la majorité sont toute sont tolérants mais ... y a toujours .. une p'tite partie qui ...

E. : bon eh ... je suis sûr que ... le Canada a d'la misère à s'accepter pour eh ... l'anglais et le français ... mais si on j'compare à d'autres pays ... comme

L. : (par rapport à un échange sur «l'immigrant modèle») bon eh un modèle d'immigrant c'est sûr eh c'est n'importe quel immigrant qui réussit à conserver sa culture ... mais en même temps s'intéresser à d'autres cultures parce que i faut pas non plus rester borné

(iii) exemples d'atténuation modale

Cette opération est omniprésente et caractérisée par une très grande variété de marqueurs : adverbess, auxiliants, parenthétiques, modes spécifiques, formes épistémiques ou aléthiques, etc. Il semblerait que cette classe de marqueurs transcatégoriels présente certains traits communs d'une langue à l'autre (Chambon & Simeoni 1997).

O. : ben eh ché po ... j'ai juste eh ... Canadien j'pense mais .. t'sé c'parce que c'est ici que j'ai vécu le plus eh

R. : J'dirais pas vraiment toutes pareilles moi les langues c'est comme pas vraiment une chose importante t'sé

3.3.2.1.2 L'opération de *remplissage* procède de même de façons fort diverses, certaines stratégies recoupant celles de la procédure de *vidage* :

(i) exemples d'atténuation et d'euphémisation

C. : ben moi j'pense que je suis un peu des deux [...] je crois que je suis un peu des deux [...] j'crois que ... j'me sens entre les deux

P. : là ma mère elle est est plus eh genre eh ben t'sé Malienne a vient du Mali là

R. : ben j'dirais qu'au primaire ça été dur

L. : j'trouve que c'est qu' surtout dans les médias ou encore eh enc dans les stéréotypes on voit surtout les immigrants comme un peu un inconvénient ou négativement

(ii) Le *remplissage* peut parfois exiger beaucoup de circomlocutions, marquant ce que nous avons précédemment analysé comme *difficulté d'énoncer* :

L. : bon disons que c'est c'est un peu dur à déterminer parce que .. en mon si j'parle dans mon cas là c'est c'est tellement ... c comme eh ... incrusté dans ma personne f que j'peux pas distinguer tout ça c'est Québécois ça c'est slovène

(iii) exemple de réplique mêlant en alternance *vidage* et *remplissage* :

E. : OK. c'parce que ... moi je me considère plus Africain que canadien parce que ben ... c'est sûr que j'pourrais pas dire ça ... vraiment sûr ... mais parce que j'ai ... toujours vécu ... plus souvent ici j'suis rarement allé en Afrique pi eh ... j'ai plus eh ... un regard sur eh le Canada comme eh ... t'sé j-j'ai grandi ici mais eh ... comme ... comme je suis eh ... chu pas blanc ... chu pas de couleur blanche mais j'peux pas dire vraiment que .. j'appartiens plus ... au ... au Canada pour moi ben j'appartiens plus à l'Afrique

Il s'agit bien d'une stratégie globale de construction du sens (Fall & Simeoni 1995) qui doit «faire avec» différents **ordres de mémoire** : ce qui a déjà été dit dans le cadre de l'échange discursif (mémoire courte permettant la cohésion du discours); ce que le locuteur pressent être le(s) point(s) de vue des autres participants (mémoire longue attribuée à autrui ou au groupe); ce que le locuteur porte en lui comme charge mémorielle (le plus souvent à son insu) et qui va se retraduire par une configuration de marqueurs en apparence idiosyncratique. Il importe de garder à l'esprit que ce qui est idiosyncratique n'est que le contenu particulier affecté aux configurations (ou constellations) formelles.

Celles-ci paraissent au contraire, nous pensons le montrer dans notre travail, hautement régulées et en nombre somme toute moindre qu'on ne le supposerait de prime abord, s'agissant d'un objet-parole que l'on dit infiniment varié. C'est en ce sens que l'on peut parler, à notre avis, d'un système de la parole distinct de celui de la langue.

3.3.2.2 Rethématisation : les nouveaux objets identifiés le sont souvent sous forme analogique (l'usage du modalisateur «comme» est fréquent). La base de cette rethématisation peut être un élément antérieur du discours (lui-même anaphorisé) ou un «savoir partagé» implicite, réactualisé :

-N : ... c'est juste la couleur de la peau qui change ou la manière d'être mais ça fait pas de différence vraiment

-Question : Est-ce que t'est dans la manière d'être qu'est-ce que tu eh tu tu entends par là manière d'être ?

-N : Ben comment qu'i peuvent penser ou agir des fois là

3.3.2.3 Pouvoirs de la rethématisation. Invités à fournir des exemples pour étayer leurs recatégorisations (ou spécifications) notionnelles, les informants s'avèrent presque toujours incapables de le faire.

-Question : peux-tu nous parler de cette différence là cette eh qu'est-ce que c'est quoi qu fait la différence

-M. : ben je l'sais pas mais ... ben ... je l'sais pas du tout

•

-Question : est-ce que t'as un exemple ... dans ta tête comme ça est-ce que tu te souviens de quelque chose que t'as entendu ?

-A. : ben ... j'ai pas vraiment d'exemple comme ça là vraiment qui ...

•

-Question : mais ou mais est-ce que tu as des exemples que tu pourrais nous donner de te dire eh

-N. : non

•

-Question : mais toi tu sais pas eh t'as pas d'idée eh

-N. : ben j'ai pas vraiment d'exemple là mais t'sé ... comment qu'i pensent pi leur culture là

Tout se passe comme si, dans l'ordre de la parole, l'opération de rethématisation était une exigence facile à remplir — les locuteurs ne semblant pas en mesure de se satisfaire d'une instabilité notionnelle durable — tandis que l'exemplification pose problème. Nous ne pouvons guère, sur ce point, que formuler des hypothèses. Est-ce parce que la rethématisation est plus sécurisante, au contraire de l'exemplification plus inquiétante ou «risquée» car susceptible d'ouvrir un nouveau front dans le jeu de contestation de frontières ? La rethématisation passe-t-elle davantage par une mémoire partagée, doxologique et donc plus directement accessible ? Cognitivement, doit-on faire l'hypothèse de noeuds plus éloignés ou hiérarchisés différemment dans le réseau des mémoires activées dans le cas de l'exemplification ? L'explication est-elle structurelle ou conjoncturelle (les circonstances, potentiellement conflictuelles, empêchant la mise à plat

des notions sensibles) ? L'interprétation de ce genre de données est clairement sujette à spéculation et nous ne pensons pas, en l'état actuel de notre recherche et des travaux dont nous avons connaissance, être en mesure de trancher.

4. IMPLICATIONS

Les formes relevées sont cumulatives. Individuellement et considérées dans ce qui les distingue de leur voisinage immédiat, elles apparaissent marginales et pour tout dire *inclassables*. Le regard objectivant du linguiste — empiriciste ou structuraliste — est donc longtemps passé à côté d'elles, les abandonnant aux stylisticiens, ou aux critiques littéraires formés à la philologie (Spitzer 1970). *Ensemble* pourtant, ces formes contribuent pleinement à la redéfinition des domaines notionnels en négociation et, au delà, au sens des mots. Il y a donc bien là les traces d'un travail mémoriel en perpétuel devenir. Chaque marque de modalisation permet ainsi d'infléchir le sens «donné» des lexèmes de la langue. Il n'est pas interdit de penser que l'échange conversationnel courant joue un rôle prépondérant dans l'évolution linguistique. Or le principe saussurien de linéarité syntagmatique interdit de penser cette évolution.

Le principe de reconstruction référentielle étant extérieur à la langue, on doit bien le rattacher à quelque chose. Selon nous, l'explication réside souvent dans les enjeux socio-culturels mis en oeuvre dans l'interaction, enjeux analysables sociologiquement. L'ancrage étant le sujet énonciateur, il existe une forte probabilité théorique pour qu'un lien existe avec *l'habitus social* du sujet (le cas échéant, son *habitus* spécifique si l'échange met en contact des pairs au cours d'une discussion professionnelle). Rappelons que *l'habitus* a une double fonction à la fois structurée et structurante. La première dimension est liée au premier temps mnémophorique; nous acquérons des façons de dire qui sont autant de façons de penser, de voir et de distinguer les choses sociales entre elles. La deuxième est à mettre en relation avec le deuxième temps dont nous avons parlé ici; nous nous évertuons à imposer nos façons de dire à autrui.

L'analyse ainsi menée devient alors éclairante pour les domaines connexes. Ainsi la traduction peut utilement se concevoir comme une ré-énonciation (Mossop 1983; Folkart 1991) exemplaire engageant non une mythique *réplication* — tropes de la «fidélité» et de son contraire, la «trahison» — mais des *substitutions* mémorielles. Il y a en effet nécessairement perte de sens (Robin 1993)¹ mais aussi gains (Cronin 1995) — non seulement substantiels mais structurels, au sens non de la structure de la langue, mais de celle de la parole et de l'écrit.

La traduction devient du coup le paradigme de tout échange énonciatif intralingual, ce que l'herméneutique a toujours su, et que certaines langues gardent en mémoire dans la façon dont elles réfèrent à l'opération de traduction/interprétation/compréhension. Le conflit mémoriel qu'actualise de façon spectaculaire la traduction serait emblématique de celui qui parcourt tout échange conversationnel : même travail de ré-énonciation, même

¹ La traduction en effet ménage bien des pièges. Il s'agit d'un exercice paradoxal et bien sûr impossible : faire passer d'une langue à l'autre *tout le système* des connotations culturelles que portent les mots, *toute la mémoire* cristallisée dans les syntagmes, les images, le lexique, *toute* la «vision du monde» portée par la langue.

(«Langue-délire et langue-délict», *Discours social*, 1993, vol. 5, 3-4, p. 9. Nous soulignons.)

effort de compréhension mutuelle (ou de manipulation), mêmes risques de malentendu, mêmes enjeux paralinguistiques (ou «métalinguistiques» au sens Bakhtinien). Le temps dédoublé des structures mnémophoriques actualisées par la parole renvoie ainsi, à terme, au mythe Babélien : chaque échange — et dans l'ordre de l'écriture chaque écrit — entreprend de reconstruire la tour, cahin-caha, avec plus ou moins de bonheur, sans que soit jamais acquise la visée stabilisatrice. L'étude du mouvement des mots qu'appelle de ses vœux Louis Lambert (Balzac, 1968 [1842]) serait alors celle des traces de la nostalgie d'un consensus impossible.

RÉFÉRENCES

- BALLY, Charles (1951) : *Linguistique générale et linguistique française*, coll. «Francke», Berne, 440 p.
- BALZAC, Honoré de (1968 [1842]) : *Louis Lambert*, Paris, Galimard.
- BOURDIEU, Pierre (1989) : *La noblesse d'état : grandes écoles et esprit de corps*, coll. «Le sens commun», Paris, Éditions de Minuit, 568 p.
- BOURDIEU, Pierre (1994) : *Raisons pratiques : sur la théorie de l'action*, Paris, Éditions du Seuil, 251 p.
- CHAMBON, Adrienne S. & Daniel SIMÉONI (en impression) : *Modalisation in the Therapeutic Dialogue*, coll. «Studies in Discours and Grammar : Linguistic Choice Across Genres : Variation in Spoken and Written English», Amsterdam et Philadelphie, John Benjamins Publications.
- CRONIN, Michael (1995) : «Shoring up the Translator's Discourse : Complexity, Incompleteness and Integration», *Meta*, Presse de l'Université de Montréal, vol. xl, 3, pp. 359-366.
- CULIOLI, Antoine (1984) : «Cours de DEA», Département de recherches linguistiques, Université Paris VII, 202 p.
- FALL, Khadiyathoulah, SIMÉONI Daniel & Adrienne S. CHAMBON (1994) : «Intégration des uns, intégration des autres. Une approche lexico-discursive de la notion d'[intégration] en situation d'interaction», K. Fall, D. Siméoni & G. Vigneaux (dir), *Mots/ Représentations. Enjeux dans les contacts interethniques et interculturels*, coll. «Actexpress», Ottawa, Presses de l'Université d'Ottawa, 418 p.
- FALL, Khadiyathoulah & Daniel SIMÉONI (1995) : «Le sens en construction», *Discours social / Social Discourse*, Montréal, vol. 7, n° 1-2.
- FOLKART, Barbara (1991) : *Le conflit des énonciations : Traduction et discours rapporté*, coll. «L'Univers des discours», Candiac (Québec), Les Éditions Balzac, 481 p.
- FREI, Henri (1929) : *La grammaire des fautes*, Paris, P. Geuther, 317 p.
- LYONS, John (1982) : «Deixis and Subjectivity : Loquor, ergo sum ?», R.J. Jarvella et W. Klein (dir.), *Speech, Place and Action*.

- MOSSOP, Brian (1993) : «The translator as rapporteur : a concept for training and self-improvement», *Meta*, Presse de l'Université de Montréal, vol. 28, n° 3, pp. 243-278.
- ROBIN, Régine (1993) : «Langue-délire et langue-délict», *Discours social / Social Discourse*, vol. 5, n° 3-4.
- SAUSSURE, Ferdinand de (1982) : *Cours de linguistique générale*, Paris, Payot, 331 p.
- SIMÉONI Daniel, FALL Khadiyathoulah & Marteen BUYCK (1994) : «Tâtonnements énonciatifs et appropriation notionnelle», J. Boucaren, J.-J. Franckel & S. Robert (dir). *Langue et langage. Problèmes et raisonnement en linguistique. Mélanges offerts à Antoine Culioli*, Paris, Presses Universitaires de France, pp. 405-411.
- SIMÉONI, Daniel & Khadiyatoulah FALL (1993) : «Bafouillage et appropriation thématique : lieux de négociation et de conflit dans l'énonciation en situation d'entretien», *Revue québécoise de linguistique*, vol. 22, 1, pp. 203-239.
- SPITZER, Léo (1970) : *Études de style*, Paris, Gallimard, 535 p.

LA MÉMOIRE CINÉTIQUE DES TERMES SUPPORTS

Amr Helmy IBRAHIM

Université de Franche-Comté, Besançon, France

1. BEAUCOUP DE MOTS SONT DES *MOBILES*

Dans *En lisant, en écrivant*, Julien Gracq fait cette réflexion :

«Même dans la prose, il faut que l'on sache tenir tête au sens. On n'est pas écrivain sans avoir le sentiment que le son, dans le mot, vient lester le sens, et que le poids dont il est ainsi doté peut l'entraîner légitimement, à l'occasion, dans de singulières excursions centrifuges. L'écriture comme la lecture est mouvement, et le mot s'y comporte comme un mobile dont la masse, à si peu qu'elle se réduise, ne peut jamais être tenue pour nulle, et peut sensiblement infléchir la direction».¹

Ce que l'intuition de l'écrivain perçoit ici très justement et à quoi il donne une aussi belle formulation stylistique, le linguiste se doit d'en élucider le fonctionnement et de montrer, si possible par une procédure reproductible, par quel côté l'effet artistique particulier à l'œuvre ici rejoint des lois plus générales qui gouvernent l'ensemble des langues.

L'idée de Gracq se laisse facilement reformuler ainsi : *quel que soit l'usage que l'on en fait en discours, les mots gardent toujours la mémoire de leur forme lorsqu'ils se combinent à d'autres mots pour faire sens. Cette forme a une inertie propre, qui peut être relativement indépendante de leur sens généralement admis. Cette inertie, ce poids, leur impriment un mouvement et peuvent les pousser à aller dans des directions qui vont infléchir l'orientation sémantique générale du discours. Elle équivaut également à dire que les mots sont, par leur seule forme, des mobiles susceptibles de déstabiliser ou tout au moins de perturber l'isotopie du texte.*

Ce mouvement virtuel ce cinétisme potentiel dont nous gardons une trace formelle dans la morphologie du signifiant mais qui n'apparaît que de manière irrégulière et aléatoire dans l'interprétation la plus immédiate et la plus conventionnelle du signifié est probablement à l'origine de l'un des principaux facteurs de diversification de l'imaginaire des lecteurs. Il a préoccupé quelques linguistes dans une perspective beaucoup moins

¹ Paris, José Corti, 1981, p. 148.

onirique. Notamment ceux qui, comme Gustave Guillaume, ne pouvaient pas concevoir qu'une bonne explication en synchronie ne soit pas *aussi et dans les mêmes termes* une bonne explication en diachronie. Son interprétation de l'évolution du *système* des oppositions des formes verbales du français en partant d'un sous-système morphologique beaucoup plus élémentaire opposant un petit nombre de terminaisons, principalement *ai* vs *ais* et *présence* vs *absence* du *r* est, de ce point de vue, exemplaire.

1.1 L'effet de *rémanence*

D'une manière plus générale et indépendamment de la question spécifique du *cinétisme* on peut dire que la réalisation d'une forme linguistique à laquelle on souhaite donner un maximum d'impact se moulera sur une représentation plus ou moins précise, un *souvenir morpho-phonique* qui peut être objectivement faux c'est-à-dire historiquement en synchroniquement — dans le système actualisé de la langue — injustifié, comme c'est le cas dans bon nombre d'étymologies dites populaires, mais qui n'en est pas moins prégnant au point d'orienter l'évolution d'un des *micro-systèmes* de la langue, voire de constituer une *classe naturelle*, c'est-à-dire un ensemble où un certain nombre de propriétés morpho-syntaxiques et un nombre significatif de traits sémantiques suivent une évolution parallèle. Nous avons appelé cet *effet de la mémoire linguistique* un *effet de rémanence* (Ibrahim 1993).

Dans le cas particulier de l'emploi comme support d'un terme la rémanence dans l'emploi support par rapport à l'emploi distributionnel de base est d'ordre *cinétique*. Concrètement cela signifie que lorsque des termes comme *passer, tirer, prendre, coup, avec, donner, pour, style, genre, dans*, se grammaticalisent et perdent tout ou partie du sémantisme qu'ils ont dans leur emploi lexical plein qui est aussi leur emploi combinatoire libre également appelé emploi distributionnel il leur reste de manière plus ou moins réelle ou virtuelle des traits sémantiques de mouvement (déplacement orienté ou non, directionnalité, itération, interruption, reprise, mouvement circulaire ou semi-circulaire, alternance, immobilisation, vitesse, ralentissement, accélération, mouvement désordonné) qui ont un rôle déterminant dans ce que j'appelle *l'actualisation métonymique d'un mot pour l'introduire dans le discours* (Ibrahim 1996a & 1997). Ce qui reste aux supports tels que nous les définissons c'est-à-dire *des lexèmes ou fractions de lexèmes vides ou semi-vides sélectionnés par un item lexicalement plein qu'ils introduisent dans le discours en en manifestant un ou plusieurs traits métonymiques, c'est une dynamique cinétique potentielle dont la combinaison avec un item lexical plein parachève le sens de cet item pour un emploi précis et le rend définissable de manière univoque*. Cette dynamique cinétique recouvre en partie mais non totalement ce que l'on appelle en allemand *aktionsart* et qu'on peut traduire en français par *mode d'action* à ne pas confondre avec l'aspect. Le mode d'action fait partie du sémantisme intrinsèque d'un item. Par exemple *contourner* a un mode d'action spécifique (tourner partiellement autour en suivant une certaine direction) quelle que soit la forme verbale qui l'actualise. Il peut avoir un aspect accompli dans *il a contourné* ou inaccompli dans *il contourna*. Le mode d'action est inscrit dans le lexique, l'aspect est le produit d'un opérateur grammatical. La dynamique cinétique potentielle est une forme dégradée et, comme son nom l'indique, pas forcément réalisée, du mode d'action.

1.2 Une fonction définitoire

Parallèlement, les nombreuses propriétés dégagées au cours des vingt dernières années à propos des différents types de supports — le premier travail approfondi sur les verbes supports est la thèse d'Anne Daladier en 1978 — nous ont amené à leur assigner une *fonction définitoire*. La définition d'un terme comme *tribunal* peut en effet être le produit de la convergence d'un certain nombre de matrices définitoires dont l'une peut avoir la forme :

Un tribunal est un lieu (où + devant lequel) comparait quelqu'un qui passe en jugement.

dans la mesure où contrairement à *juger* qui est très général, très polysémique et peut relever de plusieurs domaines, *passer en jugement* ne concerne que le domaine strictement juridique.

Ce qui reste de l'une des valeurs lexicales pleines de *passer* a gardé, dans ce contexte définitoire du support *passer* une partie de la mémoire cinétique du terme (un déplacement indéterminé par rapport à un point virtuel avec une pause devant le terme à définir, en l'occurrence *tribunal*). Ainsi les supports sont-ils souvent, du fait de cette mémoire cinétique, l'une des traces morphologiques et lexicales principales d'une partie des mécanismes d'énonciation. Ils matérialisent dans la représentation sémantique la nature du mouvement nécessaire à la conceptualisation de la relation entre l'énonciateur et les éléments qu'il place dans son discours.

1.3 Un petit exemple

Si maintenant l'on compare :

- (1) Elle prend tous les crayons qu'elle trouve.
- (2) Elle a pris le livre sur l'étagère et l'a posé sur mon bureau.
- (3) On ne se connaissait que depuis une demi-heure que déjà elle me prenait dans ses bras.
- (4) Elle m'a pris à bras-le-corps et m'a jeté dans la piscine.
- (5) «Elle prenait mon bras et nous marchions sous les arbres.» (Fromentin - TLF)
- (6) Et la voilà qui prend la tête du cortège sous les yeux médusés du Tout-Paris syndical.
- (7) Tes concitoyens universitaires ? Ils sont toujours là pour prendre le pouvoir.
- (8) Mais pour prendre leurs responsabilités,
- (9) ou tout simplement une décision c'est une autre affaire.
- (10) Il ne peut pas voter sans avoir pris part à tous les débats.
- (11) Méfie-toi, à la première odeur de conflit, il prendra la tangente.

On peut considérer que dans (1), (2), (3) & (4) on a l'ensemble des traits sémantiques d'un emploi distributionnel transitif de *prendre* à savoir :

- (a) z effectue un mouvement dans la direction de x ;
- (b) z entre en contact avec x au moyen d'une partie de son corps en saisissant tout ou partie de x ;
- (c) z déplace l'intégralité de x dans une certaine direction.

(d) Si la direction n'est pas précisée explicitement le déplacement s'effectue en direction de z .

Avec (5) on vérifie (a) et (b) mais pas du tout ou faiblement (c) et (d).

Avec (6) on vérifie (a) & (c) mais ni (b) ni (d).

Avec (7), à condition de prendre au pied de la lettre l'expression *se saisir du pouvoir*, on vérifie (a) et (b) de manière métaphorique; et, à condition de prendre au pied de la lettre les expressions *le pouvoir lui est revenu* ou *le pouvoir est repassé entre ses mains*, on vérifie également (c) & (d).

Avec (8) l'équivalence avec *assumer ses responsabilités* vérifie via une métaphore lointaine (b) et (d) mais il n'y a plus trace de (a) ou de (c).

Avec (9) on a un cinétisme en creux par la seule métaphorisation de (b) dont l'équivalence *arrêter une décision* porte la trace. Rien ne permet de vérifier (a), (c) ou (d).

Dans (10) l'équivalence avec *participer* qui présuppose de *faire mouvement vers*, voire *d'entrer dans*, permet de vérifier (a) et lui seul.

Enfin avec (11) l'expression garde certes la mémoire, très forte, d'un mouvement mais il s'agit d'une mémoire beaucoup plus holistique dont les composants sont démotivés, y compris du point de vue de leur mémoire cinétique historique, au sens où ils n'entrent plus dans des réseaux de transformations paraphrastiques et n'appartiennent plus à des paradigmes dont tous les éléments peuvent être actualisés dans la même position syntaxique — tangente n'a pas dans *prendre la tangente* de paradigme. Et c'est assurément, l'une des raisons qui feront qu'on parlera alors d'expression figée plutôt que d'expression à support.

On remarquera d'autre part que si dans beaucoup de langues la traduction donnera, à quelques exceptions près, pour (1), (2), (3) & (4) un même équivalent de *prendre* ce n'est pas le cas des exemples suivants à l'exception de (9) où l'on retrouve souvent un équivalent exact ou dérivé de *prendre*.

En d'autres termes, nous avons affaire à une échelle de déperdition du cinétisme en allant de l'emploi libre à l'emploi figé et cette déperdition entraîne une déperdition et une opacification du sens, c'est-à-dire qu'il devient de plus en plus difficile d'assigner une représentation sémantique claire et reproductible à *prendre*. Ce phénomène a souvent été remarqué pour le passage d'un emploi distributionnel à un emploi figé. Ce que nous soulignons de nouveau ici c'est que l'on peut à travers les emplois supports suivre le cheminement, parfois fort complexe, de cette déperdition qui est aussi, et c'est là une source essentielle pour l'étude de l'acquisition comme pour celle de l'évolution du langage l'un des témoins privilégiés de la reconstitution de la mémoire linguistique toutes les fois qu'une nouvelle actualisation reconnue par un nombre significatif de locuteurs acquiert un statut stable et modifie l'équilibre des configurations existantes.

2. PASSER PLUTÔT QUE FAIRE

Le fait que les supports aient d'abord été étudiés en français à partir de *faire* se justifiait statistiquement et dérivait naturellement du fait que ce sont les observations de Zellig S. Harris sur une classe particulière d'*opérateurs* qui a conduit à l'étude des supports. Il est vrai d'autre part que si l'un des exemples de départ les plus couramment cités proposés par Z. S. Harris *He walked for ten minutes / He took a ten minutes walk* se traduit en français avec *faire*, les verbes anglais qui pourraient, en dehors de cet exemple, être le plus souvent des équivalents de *faire* : *to make* et *to do* n'appartenaient pas à la

classe d'opérateurs visée par Z. S. Harris. Ils ont certes des emplois de *light verbs* mais ce ne sont pas, loin de là, les plus représentatifs ou les plus significatifs de cet ensemble. En fait, *faire* masque par son fonctionnement les propriétés les plus intéressantes des verbes supports. Non seulement parce que son cinétisme dynamique est quasiment nul mais aussi parce qu'il est le plus souvent un support générique et non un support spécifique du terme qu'il actualise, c'est-à-dire qu'il fonctionne souvent comme substitut d'un support plutôt que comme support.

Passer pourrait être le prototype des supports spécifiques. Si l'on part de cette série d'emplois distributionnels :

- 12) Passe-moi le sel [Emploi transitif de structure identique à *donner* N° -hum V *Det* N^I à N^2 -hum]
- 13) «Certaines naïvetés passent en anglais qui ne sont pas supportables en français» (J. Green - TLF) [N° -nr V en N -nr]
- 14) «Vous faites une préparation de crème ainsi qu'il suit : passez dans une casserole, avec 2 onces de beurre frais de Gournay ou d'Isigny selon la saison, quelques champignons...» (La Reynière - TLF) [N° -hum V dans *Det* N -nr]

On peut dégager les traits suivants, assignables dans une représentation sémantique : (a) *déplacement d'un élément par rapport à un point fixe concerné par le déplacement.* (b) *Aucune indication du point de départ ou du point d'arrivée de l'élément* (c) *Le mouvement se focalise sur un point du trajet du déplacement ou sur le simple fait d'effectuer ce trajet.*

Si on observe maintenant les énoncés suivants :

- 15) Je lui ai (passé + donné) un coup de fil hier.
- 16) «Je passe mille amitiés à tout le monde. Donne de mes nouvelles à notre bon grand-père». (Stendhal - TLF)
- 17) Tu (passeras + *donneras) le bonjour à Anita.
- 18) Nous avons (passé + *donné + #fait) une annonce.
- 19) Nous avons (passé + #donné + #fait) (un test + une épreuve + un examen).
- 20) Nous avons (passé + fait + *donné) la visite médicale.
- 21) «Sa mère lui passait toutes ses volontés.» (A. Fournier - TLF) [=satisfaire]
- 22) «On lui passait son château, ses équipages et sa maîtresse parce qu'il savait parler patois, acceptait volontiers une pipe de belge et appréciait le genièvre de Wambrechies» (Van der Meersch - TLF) / Il faut lui passer ses fautes. [=excuser]
- 23) «Ce Grandet ne dépense pas 10 louis sans penser à la position qu'il occupe dans le monde. Ni lui, ni sa femme ne se permettent *les caprices que je me passe*, moi fils de famille.» (Stendhal - TLF) [=que je me permets de m'offrir]
- 24) «Oui, mes enfants, je vous en passe un papier que ça serait plus pareil si les Patriotes avaient gagné en 37.» Emploi propre au Canada selon TLF / Jean a endossé un (chèque + effet + traite) / Jean a transmis la propriété d'un (chèque + effet + traite) à Georgette par endossement / Jean a *passé* un (billet + lettre de change) à l'ordre de Georgette en inscrivant l'ordre de transfert au dos de cet effet de commerce c'est-à-dire en *portant* au dos de ce titre de crédit ou de cet effet de commerce l'ordre de le payer à une tierce personne.
- 25) «Un amateur qui *passé professionnel* se lance dans des épreuves plus sévères.» (TLF) / La communauté attend un livre majeur de qui veut passer maître.

- 26) La crainte de *passer sous domination* étrangère détermina les Louisianais à s'assembler. (TLF)
- 27) «Quand les beaux jours sont arrivés, je suis venu ici (...) Le temps a passé avec une rapidité prodigieuse; je n'en ai jamais *passé* aucun qui m'ait paru plus agréable.» (A. de Tocqueville à Gobineau - 1855 - TLF) / Je ne chante pas pour *passer* le temps / Il *passa son temps* à dire du mal des gens / Il a *passé* un temps fou à réparer la chaudière.
- 28) Là-bas ? L'horreur! Là-bas c'est l'horreur Ce qui *se passe* là-bas c'est l'horreur Il *se passe* des choses horribles là-bas Que *se passe*-t-il ? Il s'en *passa* de belles!

On constate que l'interprétation de *passer* et singulièrement de la séquence nécessaire à l'interprétation de *passer* présente de manière plus ou moins nette et selon une gradation à peu de chose près croissante de (15) à (28), les caractéristiques suivantes :

- I/** Il n'est plus possible d'assigner à la séquence pertinente pour interpréter *passer* une représentation sémantique qui tienne compte de la réunion des traits (a), (b) et (c) qui sont devenus partiellement ou totalement non pertinents sinon faux.
- II/** Les nouvelles interprétations ne sont pas néanmoins totalement coupées des anciennes et connaissent des équivalences ou des incompatibilités comparables avec d'autres verbes supports comme *donner* ou *faire*. Mais ce qui est commun aux deux séries d'emplois est beaucoup plus général. Par exemple entre (12) et (15) il y a bien *quelque chose qui passe de quelqu'un à quelqu'un d'autre* mais alors qu'en (12) le verbe est un opérateur dont les trois arguments explicites (quelqu'un, le sel, quelqu'un) définissent simultanément une structure et une représentation sémantique isomorphe à cette structure, en (15) le verbe n'est plus qu'un actualisateur, un élément qui permet à *coup de fil* d'assumer une fonction analogue à celle d'un opérateur avec cette conséquence que l'énoncé ne dit plus rien, ni explicitement ni implicitement, sur la nature de ce qui circule entre les deux *quelqu'un*. C'est un savoir extralinguistique très différent de celui dont nous avons besoin pour comprendre *sel* qui est nécessaire pour reconstituer à partir de *coup de fil* une structure syntactico-sémantique sous-jacente effacée du type : *ensemble de mots constitués en énoncés produits oralement par quelqu'un et transmis par l'intermédiaire d'un appareil qu'on manipule avec des mouvements dont certains présentent des analogies, notamment lorsque l'on appuie, décroche ou raccroche, avec les mouvements qui accompagnent les coups, et qui circulent à l'intérieur de fils à destination de quelqu'un qui les reçoit et peut y répondre par l'intermédiaire d'un appareil analogue*. Alors que l'interprétation de *sel* repose sur la simple association d'un signe et d'une image référentielle, celle de la structure où intervient *coup de fil* est le résultat d'une succession de *réductions* rendues possibles par le parallélisme du caractère métonymique tout à la fois de la représentation sémantique et de la structuration syntaxique : la structure définitoire analytique étendue ne comporte aucune position ni aucun terme qui ne puissent pas être réduits selon les mécanismes généraux de la langue à l'énoncé synthétique de départ.
- III/** Dans (21), (22) et (23) le sens de la séquence comportant *passer* avec les valeurs de *satisfaire*, *excuser* ou *se permettre* est entièrement construit sur une *partie* de la

propriété cinétique de *passer* : un passage que l'on peut *laisser se faire* ou au contraire *entraver* ou *arrêter*. Il vient de la possibilité de greffer une métaphore sur cette propriété métonymique, accroissant considérablement la distance qui sépare la valeur support de la valeur distributionnelle de la première série d'énoncés.

- /IV/ En (24) on peut voir de manière exemplaire comment la mémoire du support est consignée aussi bien dans une expression familière — l'expression canadienne — que dans une langue de spécialité — en l'occurrence le langage juridico-financier. Les définitions matricielles que nous donnons sont à la fois des définitions lexicographiques que l'on peut relever sous une forme plus ou moins complète dans différents dictionnaires et, pour partie, des *expressions* de la langue actualisée. On voit aussi que la formulation la plus synthétique et la plus éloignée de la valeur distributionnelle est, comme on peut s'y attendre, l'expression familière équivalent à *je vous garantis*.
- /V/ (25) et (26) donnent le résultat de la circulation, du *passage* : un changement d'état. La valeur aspectuelle du cinétisme est ici dominante.
- /VI/ L'interchangeabilité *temps qui passe / passer son temps* est très caractéristique de la transformation d'un opérateur en support sans qu'il y ait un changement radical dans le sens de l'énoncé ou avec un changement qu'il est relativement difficile de représenter précisément. On peut facilement donner une représentation métaphorique du temps qui passe : le temps est l'argument d'un opérateur circulant. Par opposition *passer le temps* n'a aucun sens assignable. Il faut d'ailleurs *passer son temps à faire quelque chose* ce qui est l'équivalent exact de *faire quelque chose* avec ce rappel que *tout ce qui se fait se fait dans le temps qui passe en même temps que le temps où je suis*. Cette configuration apparemment circulaire du sens favorise les réductions. Selon la portion réduite on a le temps qui passe ou celui que l'on passe. La séquence réduite garde la mémoire de la matrice complète mais ne garde pas la totalité de ses significations. Parallèlement les différentes réductions n'ont pas le même sens.
- /VII/ Dans le dernier exemple (28), le support *passer* ou *se passer* actualise un énoncé entier et ne garde plus de la valeur distributionnelle du verbe que *la mémoire d'un mouvement pur* ou encore ce que l'on pourrait appeler cognitivement *la part de mouvement ou de cinétisme qui constitue la seule composante absolument irréductible de la représentation d'un événement*.

RÉFÉRENCES

- IBRAHIM, Amr Helmy (1993) : «La déviance de la suffixation en français est-elle structurelle ?», *BULAG-TRANEL*, n° 20, décembre, *Le traitement des données linguistiques non standard*, Actes des rencontres Besançon-Neuchâtel, Neuchâtel, 29-30 janvier 1993, Besançon, Université de Franche-Comté (France), Neuchâtel, Université de Neuchâtel (Suisse), pp. 113-125.
- IBRAHIM, Amr Helmy (1996a) : «Les supports : le terme, la notion et les approches» & «La forme d'une théorie du langage axée sur les termes supports», *Langages*, n° 121, mars, *Les supports*, Paris, Larousse, pp. 3-8 & 99-120.

- IBRAHIM, Amr Helmy (1996b) : «Peut-on, en français, reconnaître automatiquement un support de péjoration ?», *LYNX*, n^{os} 34-35, *Hommage à Jean Dubois*, Nanterre, Centre de recherches linguistiques de l'Université Paris X-Nanterre, pp. 57-77.
- IBRAHIM, Amr Helmy (1997) : «Protocole de reconnaissance des supports suivi d'un exercice en vue d'identifier un support en anglais» & «Les supports en arabe», *Revue québécoise de linguistique*, 26 (1), *Les supports : une problématique universelle*, Montréal. Université du Québec à Montréal.

MÉTAPHORES ET DIACHRONIE DANS LES SCIENCES : LE CAS DE *CODE*, *PATRIMOINE*, *SÉLECTION*

François GAUDIN

«Sociodynamiques langagières», UPRESA 6065, Université de Rouen, France

«Nous pensons à travers des mots, à l'aide de mots, et peut-être aussi contre eux; un concept, vu par le petit bout, c'est aussi un vocable.» (Schlanger 1971 : 11). Ce petit bout de la lorgnette, c'est l'endroit où les linguistes, penchés sur les écrits des scientifiques, sont postés.

Dans le développement des disciplines, dans l'effort de conceptualisation, mais aussi de marquage de territoire que constitue le développement d'un champ, un effort tout particulier est consacré à l'élaboration d'une terminologie spécifique. Or l'une des caractéristiques de ces terminologies, c'est le mouvement de reprise de termes déjà utilisés. Par là, les disciplines s'inscrivent dans une mémoire intellectuelle, parfois délibérée, parfois inconsciente.

Que ces reprises soient liées à des effets de modes, à des influences idéologiques, à des commodités de discours, elles montrent que les champs sont reliés par le langage et que la séparation entre mots, termes et notions ne rend pas compte de la réalité des faits langagiers et intellectuels.

Ces mouvements de reprise sont de type métaphorique : le sujet parlant reprend le mot, mais il ne reprend pas l'intégralité des éléments de sens qui lui étaient associés dans des interactions précédentes.

La métaphore a retenu de nombreux auteurs, souvent soucieux d'analyse littéraire. Toutefois, son caractère culturel, voire anthropologique, a peu attiré l'attention. On le trouve mis en lumière dans des travaux aussi différents que ceux de Pierre Guiraud (1978) ou George Lakoff et Mark Johnson (1985), ou, pour sortir du champ linguistique, ceux de Gilbert Durand (1992) sur les structures anthropologiques de l'imaginaire, ou de Gérard Holton (1981) sur l'imagination scientifique.

Ces travaux laissent penser que les énoncés que nous reprenons conditionnent nos façons de dire et de voir. Dans le domaine de la pensée, certains paysages intellectuels partagent, à un moment de l'histoire, les mêmes dessins métaphoriques. Ils partagent un même espace de discours.

Parler d'espace de discours, c'est prendre le risque du flou. Mais il importe de ne pas résumer ces mouvements métaphoriques à des phénomènes affectant des signes pris isolément. Les mots ne sont pas solitaires; ils n'ont de réalité qu'au sein d'énoncés. Lorsqu'elle joue un rôle dans les verbalisations intellectuelles, la métaphore conduit à se poser des questions relevant à la fois de la phraséologie et du dialogisme. En effet, le terme que l'on reprend à d'autres discours, que l'on emprunte à d'autres conceptualisations, c'est aussi le pivot d'un micro-réseau verbal, avec ses locutions toutes faites, ses cooccurrences, ses constructions typiques, etc. Reprendre un terme, c'est bien reprendre un stock d'énoncés disponibles, de schémas de pensée capitalisés et versés au pot commun de la langue, vecteur de culture. Et ces reprises qui peuvent influencer profondément nos façons de voir ne sont pas sans présenter parfois des dangers.

CODE

Prenons le mot *code*. Tel qu'il est utilisé en génétique, il peut sembler anodin. Le mot en lui-même est d'un intérêt modeste; ce qui est remarquable, c'est qu'il véhicule avec lui un champ lexical repris du vocabulaire de l'information et de la communication : *information, transcription, traduction, message, lecture, décodage, copie*, etc. Cela suscite des façons de dire très marquées, ce qui s'explique historiquement par l'influence de la théorie de l'information. Un exemple entre mille, on lit : «L'information génétique portée par l'ARN messenger est lue et traduite en une séquence d'acides aminés» (*Atlas de poche de génétique*, Flammarion, p. 391).

L'introduction de ces concepts et de ces termes dans les années cinquante a refondé la génétique. Cette transformation terminologique a vu le passage d'une conception en termes de détermination à une vision fondée sur la notion de programme. Le vocabulaire employé ici est devenu aujourd'hui canonique, son caractère métaphorique n'est plus senti; toutefois l'analogie a conservé sa valeur heuristique. En effet, la langue permet d'établir un pont entre les séquences d'acides aminés et les langues écrites, car le mot *code* s'utilise pour les deux. Pour les biométriciens, les séquences génétiques sont composées de «lettres» et l'on y recherche l'apparition de «mots rares». Dans le cadre de cette analogie, on rencontre des énoncés tels que : «L'ajout d'une lettre dans une des séquences est biologiquement compréhensible.» (Mouchard, dans Alexandre 1995). Ailleurs, le code génétique est comparé à la «pierre de Rosette, permettant d'aller et venir entre deux "langages", celui des protéines et celui de l'ADN.» (Blocker et Salem 1994 : 74).

Mais il y a plus, l'analogie est prise au pied de la lettre et les biométriciens s'intéressent aux structures des langues pour y trouver des pistes d'investigation intéressantes pour les séquences génétiques (Pesole *et alii* 1994). Et, à tout seigneur tout honneur, on trouve même à l'occasion des évocations de l'auteur du *CLG*. Comme l'écrit Bernardino Fantini : «L'arbitraire du signe, qui avait été introduit en linguistique par Saussure, trouve en biologie sa réalisation dans la structure du code génétique, qui ne peut recevoir d'autre explication que par les hasards de l'évolution» (Fantini 1988 : 168).

Le réseau métaphorique concerné, qui a permis de rendre compatible l'approche de la biologie avec celle du traitement de l'information, permet des rapprochements avec les langues : le terme *code* permet de jouer sur trois claviers de notions à la fois celui de la reproduction de la vie, celui du programme informatique et celui des langues.

On pourrait voir là un emprunt occasionnel et de peu de conséquence. Mais ce serait méconnaître le fait que ce parallèle ne se limite pas à un effet métaphorique de surface : il influence notre conception de la réalité concernée. À parler de *code*, de *lecture*, de *transcription*, etc., il s'ensuit que la nature, sous les couverts d'un imaginaire modernisé, retrouve la figure d'un automate, donc d'un mécanisme. **La nature est alors conçue comme un agent.** On voit ici que les termes ne sont pas neutres quant à la façon dont ils déterminent nos catégorisations et notre appréhension du réel.

Il existe un autre aspect sous lequel la métaphore du code est intéressante, c'est qu'elle supporte la conception contemporaine de l'hérédité. Cette conception, largement déterministe, se retrouve dans une autre métaphore, venue du domaine financier, celle du patrimoine.

PATRIMOINE

En biologie, les cellules sont communément présentées comme possédant « un 'noyau' détenteur du patrimoine génétique sur lequel reposent la vie et la reproduction de la cellule. » (Douzou *et alii* 1987 : 12)¹. Or la notion de patrimoine n'est pas innocente. Dans l'usage le plus courant, la terme de *patrimoine* permet de reformuler la notion technique de génotype. Ce vocable, *patrimoine*, est pourvu d'une histoire et d'une charge culturelle qui oriente la pensée et tend à induire un point de vue. Le terme *génotype* désigne une notion technique, tandis que *patrimoine* propose à l'esprit une notion culturelle, car le signe est le support de représentations disponibles.

De fait, la notion véhiculée par le terme *patrimoine génétique* n'est pas innocente. Elle conduit facilement à un déterminisme héréditariste, car elle entraîne avec elle l'idée qu'il existe un ensemble constitué et transmissible de façon intangible. Cette idée est bien ancrée, même si ses bases scientifiques ont été rejetées. Comme le souligne Ernst Mayr, « la croyance en l'hérédité des caractères acquis était quasiment universelle jusqu'au début des années 1880 » (1993 : 150). Bien sûr, on sait bien aujourd'hui que le code génétique n'est porteur que de virtualités, dont l'actualisation va dépendre étroitement des facteurs environnementaux. Mais les mentalités s'inscrivent dans une durée qui n'a rien de commun avec le rythme de parution des articles scientifiques. Or l'ancienne conception fixiste entraîne avec elle une vision du génome comme ensemble d'éléments unifonctionnels et réparables; elle n'a d'ailleurs été réfutée de façon décisive que dans les années 1950 grâce aux découvertes de la biologie moléculaire.

Si certaines modifications graves condamnent à l'anomalie, concevoir les maladies génétiques comme résultant, d'une part, de l'altération d'un seul gène et, d'autre part, comme se développant indépendamment du milieu semble aujourd'hui très réducteur. Or la conception la plus répandue dans le public est nourrie par les informations relatives aux maladies dramatiques et spectaculaires, dont la mutation d'un seul gène est la cause (l'anémie drépanocytaire est due à l'altération d'une seule paire de bases dans l'un des gènes de l'hémoglobine). On s'accorde aujourd'hui (Spitz 1996 : 75) à reconnaître que de nombreuses caractéristiques du phénotype sont déterminées par l'action de plusieurs gènes, les « polygènes » (cf. Roubertoux et Carlier 1996), dont chacun a un effet mineur, mais

¹ On remarquera ici que ce n'est pas *patrimoine* qui est guillemeté, mais *noyau*, qui n'est pourtant, ici, plus guère métaphorique puisque le terme s'emploie en biologie depuis près d'un siècle et demi.

dont les effets s'additionnent. Quant aux maladies à caractère héréditaire les plus fréquentes, elles sont dues à l'influence de plusieurs mutations génétiques.

Or il faut y insister, il n'existe pas, à proprement parler, de «patrimoine génétique». La notion, devenue populaire de nos jours, de génome humain est trompeuse. Le biologiste Henri Atlan y insiste avec force : «Il n'y a que des génomes d'individus qui appartiennent à la même espèce, l'*Homo sapiens*.» (1997 : 84). C'est pourtant la vision «patrimoniale» qui triomphe et se diffuse dans le grand public. Elle est en effet au coeur de la logique de ce que l'on appelle la *thérapie génique*, dénomination trompeuse selon certains, qui préfèrent désigner cette technique par le terme de *transfert de gène*, en l'attente d'une démonstration patente de son efficacité (Laduron 1996 : 9).

Les expressions de *programme génétique*, de *patrimoine génétique* participent de deux métaphores, mais du même abus de langage : elles conduisent à penser comme transmissibles de façon mécanique des données cellulaires qui ne sont que des promesses, des virtualités. Il convient donc de se méfier de la diffusion d'une approche métaphorique qui envisage le génome comme un ensemble de données localisées sur un support matériel, l'ADN, et pourvues d'une vertu caractéristique des biens matériels, l'héritabilité.

Notre vision est donc influencée par les métaphores qui se trouvent naturalisées après leur inscription dans la culture. La rémanence de la métaphore patrimoniale constitue un facteur possible d'erreur dans l'appréhension de l'hérédité. Il nous est assez facile d'y réfléchir, car le vocabulaire de la génétique est né d'innovations sémantiques qui nous sont encore sensibles. Ce n'est plus le cas de conceptions, comme celles de Darwin, qui nous ont habitués à envisager l'évolution comme résultant d'héritages successifs et, surtout, d'une sélection que l'on dit «naturelle».

SÉLECTION NATURELLE

Quand nous pensons aujourd'hui à Darwin et à la notion de sélection naturelle, il nous est difficile d'apercevoir à quel point Darwin innovait en utilisant l'adjectif. Son coup d'éclat lexical réside dans la détermination de *sélection*, terme emprunté au langage de l'agronomie. À quoi sert ce terme en agronomie ? À désigner le fait que l'homme sélectionne les plantes en vue de leur amélioration. Avec le syntagme de Darwin, l'agent change.

Darwin lui-même explique la logique de ce changement :

«Si l'homme peut, à force de patience, choisir les variations qui lui sont utiles, pourquoi, sous l'effet de conditions de vie changeantes et complexes, des variations utiles aux produits vivants de la nature ne surgiraient-elles pas souvent et ne seraient-elles pas préservées ou choisies ?» (Darwin, *On the origin of species*, p. 445 cité dans Schlanger 1971 : 23).

On trouve ici deux schémas :

- (la volonté de) l'homme choisit les variations utiles à l'homme;
- (la volonté de) x choisit les variations utiles aux produits de la nature.

Dans les deux cas, *sélection* renvoie à la notion de «choix de variations utiles»; l'adjectif *naturelle* vient spécifier que la cause n'en est pas la volonté de l'homme, c'est

donc la «nature» qui tient la place de l'agent. Darwin le dit d'ailleurs clairement : «Je lui ai donné le nom de sélection naturelle, pour marquer sa relation avec le pouvoir humain de sélection.» (Darwin, *On the origin of species*, p. 67, cité dans Schlanger 1971 : 23). Et il est conscient des équivoques que peut provoquer son expression : «Au sens littéral sans doute, la sélection naturelle est une expression fautive» (*idem* : 24). Expression fautive, notamment, parce que Darwin ne vise pas une quelconque force ou divinité; sa vision de la nature est laïque, il ne fait que chercher un mode explicatif. En fait, chez Darwin il n'y pas d'agent sélecteur bien défini: cognitivement un actant nous manque qui prendra la forme linguistique de l'adjectif *naturelle*.

Nous pouvons préciser notre schéma, en insistant sur le parallèle :

- (la volonté de) l'homme choisit les variations utiles à l'homme;
- (la volonté de) la nature choisit les variations utiles aux produits de la nature.

Plus précisément, la sélection naturelle est une sélection compétitive, dans la mesure où Darwin l'explique par la lutte pour la vie. Ce qui permet le choix, ce n'est pas tant la nature qu'un principe. Ce principe, Darwin le dénomme en reprenant à l'économiste conservateur Malthus la notion de lutte pour la vie. Cet arrière-plan importe, dans une période où l'on «blâme les lois d'assistance publique, car elles préservent le faible aux dépens du fort.» (Barlow 1985 : 132). La vision de l'économiste Malthus est pessimiste et statique : «il y a d'incessantes catastrophes, une interminable et féroce lutte pour la vie, et pourtant le monde reste essentiellement le même.» (Mayr 1993 : 112). Mais il y a plus : ce penseur des phénomènes populationnels raisonne en considérant «les riches et les pauvres comme des catégories distinctes par essence dont il n'est pas question de discuter le statut, des **espèces** [nous soulignons] pourrait-on dire, condamnées à cohabiter dans leur opposition.» (Veuille 1987b : 211).

Sous l'influence de sa lecture, en 1838, Darwin passe de l'idée d'une **concurrence entre espèces** à celle d'une **concurrence entre individus** dont la sélection porte sur la variation des patrimoines génétiques. Dans son autobiographie, il parle clairement d'organes «développés par sélection naturelle, ou survie du plus apte [...] de telle sorte que ceux qui les possèdent puissent entrer en compétition avec d'autres êtres» (Darwin 1985 : 73). Il existe donc, selon lui, une compétition **interne** à chaque espèce qui a pour effet de maintenir l'équilibre entre l'espèce et son milieu, en ne permettant la survie que des plus aptes.

La conceptualisation de Darwin rend donc solidaires les notions de «sélection» et de «lutte pour la vie» en les rendant indépendants de leurs contextes d'origine. Ce faisant, il dénature le concept de Malthus qui passe d'un discours sur la société à un discours sur la nature. Il y a eu mouvement métaphorique, mais le discours receveur a construit un nouveau réseau linguistique décrivant un nouveau concept.

On est passé pour *sélection* de :

- (la volonté de) l'homme choisit les variations utiles à l'homme;
- la lutte pour la vie choisit les variations utiles aux produits de la nature.

Le rôle de l'agent est tenu par un concept abstrait qui tient lieu de la «volonté» de la nature. C'est donc un concept matérialiste.

Par rapport à Malthus, Darwin se détache de la vision d'une nature faite à l'image de la société; par rapport à son époque, il rompt avec le fixisme religieux pour penser le sort des espèces vivantes. La pensée de Darwin résulte donc bien d'une réflexion métaphorique, mais c'est l'exclusion, ou l'oubli, de ces métaphores qui permet à la pensée de Darwin de se faire jour et d'être comprise.

Bien sûr, on réutilisera Darwin dans l'autre sens, c'est-à-dire en réappliquant ses concepts aux sociétés humaines. Mais ce sera là un simple jeu idéologique. C'est simplement que l'analogie société / nature lui préexiste et le poursuivra mais «Rien de ce qui *caractérise* le darwinisme ne se retrouvera dans quelque forme de biologisme social» (Veuille 1987b : 217). Il n'y a plus de pensée autonome autour du terme *sélection naturelle*, parce que l'idée de sélection naturelle perd toute validité dès qu'elle est (ré-) importée.

L'exemple de Darwin permet de démontrer qu'il n'y a véritablement d'emprunt heuristique que lorsque préexiste un problème. En fournissant tout à la fois des expressions et des arguments, des représentations et des modèles, le raisonnement analogique procure à la pensée un support imaginaire. C'est-à-dire que l'analogie permet la verbalisation du concept mais n'en est pas la source. «Là où c'est le cas, on a quitté le plan de la connaissance pour celui de la rhétorique.» (Schlanger 1971 : 26). Dans le cas de la sélection naturelle et de la réimportation du concept vers le social, on a affaire à un détournement rhétorique, le terme est utilisé pour le prestige du concept construit au sein d'une autre formation discursive. Il s'agit là d'un dévoiement, car on utilise la plus-value intellectuelle procurée par l'usage d'un concept scientifique.

Cet exemple permet de mettre en lumière le caractère idéologique de certaines métaphores.

CONCLUSION

Métaphores, histoire, les pages qui précèdent ne cherchaient qu'à illustrer le bien-fondé de la prise en compte de l'histoire des signes et des idées dans le cadre d'une terminologie sociale, ou *socioterminologie* (Gaudin 1993b), ouverte à l'épaisseur historique des signes par lesquels nous formulons les savoirs.

Si, depuis une dizaine d'années, le champ des études terminologiques s'est ouvert aux préoccupations sociales, si les présupposés de la discipline ont été remis en question de façon fructueuse, si la philosophie du langage est moins ignorée qu'auparavant, si la linguistique de corpus y est plus présente que par le passé, il reste cependant à faire pour que la méta-terminologie prenne corps et s'élargisse à une étude de l'expression linguistique des connaissances et des idées.

Or ce sont sans doute les liens avec l'histoire des sciences et l'épistémologie qui se sont le moins développés depuis dix ans. Il y a là une difficulté qui tient à la culture spécialisée que réclament ces champs d'étude. Mais également, les besoins sociaux peuvent difficilement s'affirmer; ces préoccupations relèvent d'une recherche désintéressée, que notre époque n'encourage guère. Elles pourraient prendre place dans le cadre d'une approche plus culturelle de la lexicographie spécialisée, mais cette orientation ne se dessine pas encore.

C'est pourquoi il importe d'étudier «la mémoire des mots», titre d'une initiative heureuse, car trop peu de ponts sont jetés entre l'histoire des langues et celle des idées. Une telle convergence ne satisfait pas seulement au goût des douceurs théoriques, elle devrait permettre de mieux comprendre la façon dont nos façons de dire conditionnent nos façons de penser. Et si la terminologie cherche depuis longtemps rectitude des dénominations, selon l'expression d'Alain Rey, sa vocation doit également la conduire à se soucier de la clarté des conceptualisations.

RÉFÉRENCES

- ATLAN, Henri *et alii* (1997) : *Savoir échanger les savoirs*, éd. Textuel, 105 p.
- BARLOW, Nora (1985) : «Annexes», Darwin, Charles (1985) [1958] *La vie d'un naturaliste à l'époque victorienne*, éd. Belin, pp. 127-157.
- BLANCKAERT, Claude (1988) : «Variations sur le darwinisme. Épistémologie et transfert», dans Louis, Pierre et Roger, Jacques (dir), *Transferts de vocabulaires dans les sciences*, éd. du CNRS, pp. 135-157.
- BLOCKER Ariel et Lionel SALEM (1994) : *L'Homme génétique*, Paris, Dunod, 320 p.
- DARWIN, Charles (1985) [1958] : *La vie d'un naturaliste à l'époque victorienne*, éd. Belin, 175p.
- DOUZOU, Pierre *et alii* (1987) : *Les biotechnologies*, coll. «Que sais-je ?», n° 2127, Paris, PUF, 127 p.
- DUNON, Dominique et Beat IMHOF (1996) : «Inflammation et cancer : les cellules "passe-muraille"», *La recherche*, n° 283, pp. 64-68.
- DURAND, Gilbert (1992) : *Les structures anthropologiques de l'imaginaire*, Paris, Dunod, 560p.
- EPELBAUM, J. (1996) : *Neuropeptides et neuromédiateurs*, 2^e éd., éd. Inserm, 306 p.
- GAMBIER, Yves et François GAUDIN (dir) (1993) : *Socioterminologie, Le langage et l'homme*, vol. XXVIII, n° 4, Bruxelles, éd. DeBoeck Université.
- GAUDIN, François (1993a) : «Socioterminologie : propos et propositions épistémologiques», Gambier, Yves et Gaudin, François (dir), *Socioterminologie, Le langage et l'homme*, vol. XXVIII, n° 4, Bruxelles, éd. DeBoeck Université, pp. 247-257.
- GAUDIN, François (1993b) : *Pour une socioterminologie : des problèmes sémantiques aux pratiques institutionnelles*, Publications de l'Université de Rouen, 254 p.
- GUIRAUD, Pierre (1978) : *Sémiologie de la sexualité*, Paris, Payot, 247 p.
- HOLTON, Gérald (1981) : *L'imagination scientifique*, Paris, Gallimard, 496 p.

- LADURON, Pierre (1996) : «Thérapie génétique et intégrisme scientifique», *La recherche*, n° 287, mai 1996, p. 9.
- LAKOFF, George et Mark JOHNSON(1985) : *Les métaphores dans la vie quotidienne*. Paris, Minuit, 254 p.
- MAYR, Ernst (1993) : *Darwin et la pensée moderne de l'évolution*, éd. Odile Jacob. 248 p.
- MOULIN, Anne-Marie (1991) : *Le dernier langage de la médecine. Histoire de l'immunologie de Pasteur au Sida*, Paris, PUF, 447 p.
- PESOLE, Graziano *et alii* (1994) : «Linguistic approaches to the analysis of sequence information», *Tibtech*, vol. 12, ed. Elsevier, pp. 401-408.
- REY, Alain (1979) : *La terminologie, noms et notions*, Paris. PUF, coll. «Que sais-je ?», n° 1780, 127 p.
- ROUBERTOUX, Pierre et Michèle CARLIER (1996) : «Le QI est-il héritable ?», *La recherche*, n° 283, pp. 70-78.
- SCHLANGER, Judith (1971) : *Les métaphores de l'organisme*, Paris, Librairie J. Vin. 269 p.
- SPITZ, Elisabeth (1996) : «Des jumeaux bien dociles». *La recherche*, n° 283, pp. 73-74.
- VEUILLE, Michel (1987a) : «Corrélation. Le concept pirate», Stengers Isabelle (dir), *D'une science à l'autre. Des concepts nomades*, Paris, Seuil, pp. 35-67.
- VEUILLE, Michel (1987b) : «Sélection naturelle», Stengers, Isabelle (dir), *D'une science à l'autre. Des concepts nomades*. Paris, Seuil, pp. 198-218.

MAILLE OU MAILLON : QUAND DES TERMINOGRAPHES NÉGLIGENT L'ÉVOLUTION DE L'USAGE

Marc VAN CAMPENHOUDT

Centre de recherche TERMISTI, Bruxelles, Belgique

1. INTRODUCTION

À juste titre, l'urgence conduit les terminologues à se soucier en priorité des sciences et techniques récentes en vue de répondre aux besoins néologiques de la langue française. Chaque vocabulaire de spécialité s'ancre pourtant dans la diachronie et plusieurs recherches ont déjà montré — comme le rappellent Éloy et Humbley (1993) — que les nouvelles terminologies connaissent une longue période de flottement avant de se stabiliser dans la durée.

On dispose de peu de données sur la manière dont les dictionnaires spécialisés contribuent ou non à l'implantation, à la normalisation, voire à la perpétuation des terminologies. À travers l'étude de la désignation de deux concepts propres au domaine nautique et se rapportant plus précisément à la chaîne d'ancre, notre objectif dans les pages qui suivent sera de décrire un processus de mise en place d'une terminologie et d'observer la manière dont l'usage est décrit par les dictionnaires de marine au cours d'une période d'un siècle et demi.

L'emploi de chaînes pour relier les ancres aux navires ne date que du début du XIX^e siècle. Auparavant, on utilisait des cordages nommés *câbles*; aussi, lorsque les progrès techniques permirent d'utiliser le métal, on eut d'abord recours à l'expression *câble-chaîne* [1] pour désigner le nouveau dispositif, avant d'opter progressivement pour le mot *chaîne* (cf. extrait [21]).

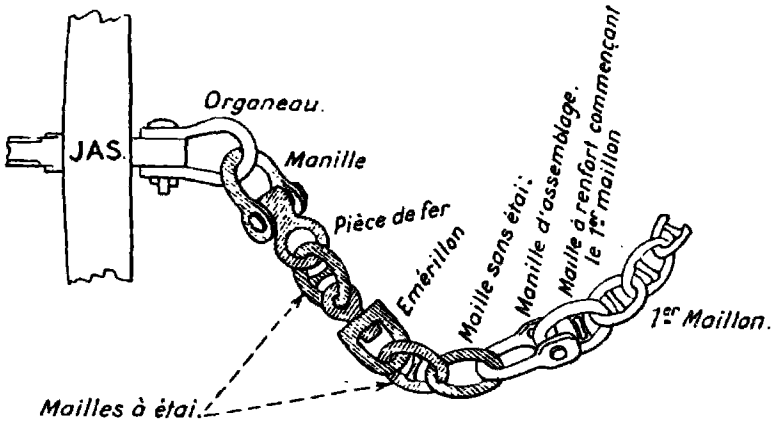
[1] «L'invention des *câbles-chaînes* est due au capitaine Samüel Brown; ses essais datent de 1808. M. Brunton les a perfectionnés en 1812. — Leur adoption est presque générale aujourd'hui.» (Lecomte 1835 : 85)

Comme le montrent les illustrations, les chaînes sont constituées d'anneaux d'une forme caractéristique. Pour la facilité de la manœuvre, elles ont la particularité d'être subdivisées en portions de 30 mètres (15 brasses en Angleterre) reliées entre elles par une manille. Les concepts qui nous occupent correspondent précisément à ces anneaux et à ces ensembles de 30 mètres. La désignation de ces méronymes a d'abord donné lieu à une large synonymie entre des mots comme *chaînon*, *maillon* et *maille*. Ensuite, l'usage a progressivement retenu deux de ces lexies comme termes monosémiques [2-3] :

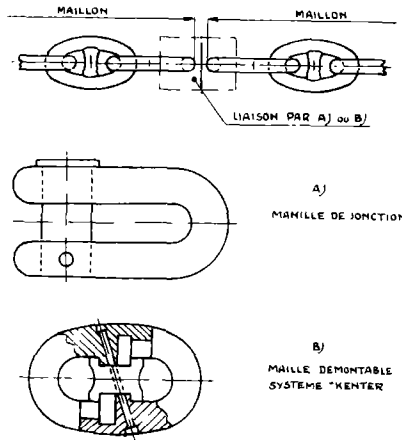
[2] «**Maille** n.f. Anneau d'une chaîne d'ancre, muni d'un étai pour éviter sa déformation.» (Sizaire 1978 : 106)

[3] «**Maillon** n.m. Élément d'une chaîne d'ancre, d'une longueur égale à 30 mètres quel que soit le calibre des mailles.» (Sizaire 1978 : 106)

Cette terminologie largement implantée au terme d'un lent processus naturel est notamment entérinée par le *Vocabulaire normalisé de la navigation maritime* (O.M.I. 1985) et par les normes ISO 3828 (1984) et 1704 (1991).



(Massenet *et al.* 1911 : I-293)



(Chicot 1960 : 293)

Tout marin qui se respecte a appris à ne pas confondre la maille et le maillon¹, comme il a appris à parler de *cordage*, de *compas* ou de *feu* plutôt que de *corde*, de *boussole* ou de *lampe*. Étrangement, cette terminologie de la chaîne d'ancre n'est pas entérinée par certains dictionnaires spécialisés de l'après-guerre. Un retour vers l'usage écrit, de 1830 jusqu'à nos jours, et une comparaison avec le contenu des dictionnaires de marine monolingues, bilingues ou multilingues parus au cours de la même période devrait permettre de clarifier la situation.

2. ÉVOLUTION DE L'USAGE DANS LES MANUELS D'ENSEIGNEMENT MARITIME

Notre enquête sur l'écrit sera essentiellement dirigée vers les manuels d'enseignement maritime dans la mesure où ces textes destinés aux élèves matelots ou aux élèves officiers ont pu jouer un rôle important dans l'établissement de l'usage. Ces ouvrages, publiés ou photocopiés, sont relativement accessibles dans les bibliothèques spécialisées et contiennent systématiquement les informations souhaitées. Par ailleurs, il y a lieu de penser qu'en vertu des mêmes arguments, ils ont pu constituer des sources d'information privilégiées pour les auteurs de dictionnaires. À défaut de constituer des marques suffisantes de l'usage réel, ils ont assurément dû exercer une influence sur les usagers; à la différence des catalogues de *shipchangers*, inventaires de bord et autres rapports de mer, par ailleurs beaucoup plus difficiles à réunir.

2.1 Une large synonymie

Dans les sources les plus anciennes, à une époque où la technologie de la chaîne constitue encore une nouveauté, on trouve une terminologie plutôt confuse, proche de la langue courante : l'élément constitutif est dénommé *maillon*² dans des instructions officielles de la marine française de 1833 [4] et dans un cours photocopié de l'École navale impériale daté de 1866-67 [5]. Dans un rapport technique de Barbotin³ (1835), ce même élément est indifféremment appelé *chaînon*, *maillon* ou *maille*, tandis que la portion de 30 mètres est dénommée *chaînon* [6-8].

[4] «L'appareil auquel on a donné le nom de LINGUET DE CHAÎNE offre les moyens de saisir la chaîne de maillon en maillon et d'annuler ainsi le contre-coup du tangage et l'effet toujours dangereux du choc de la tournevire [...]. Une rainure, qui le parcourt dans toute sa longueur, reçoit les maillons verticaux, et ceux placés horizontalement sur deux plates-bandes qui lui sont parallèles.» (*Instructions* 1833 : 3-4)

1 On notera que le terme *maille* peut également désigner l'intervalle entre deux couples de la coque d'un navire ou encore l'ouverture entre les fils d'un filet de pêche. Le terme *maillon* désigne aussi une sorte de nœud coulant utilisé pour récupérer des objets immergés. Ces acceptions sont plus anciennes que celles qui se rapportent à la chaîne d'ancre.

2 Il n'entre pas dans notre propos de nous prononcer sur le fondement historique de tel ou tel usage. Selon Wartburg (1950-1957 : II-498b et VI,1-14b et 15a), *chaînon* 'anneau d'une chaîne' est attesté depuis 1390, *maille* 'petit anneau de métal' est attesté depuis 1411 et *maillon* 'chaînon', seulement depuis 1752 (Trévoux). Nos extraits permettent, par ailleurs, de proposer des dates d'attestation antérieures à celles de Wartburg (*ibid.*) pour le domaine maritime.

3 Le capitaine de vaisseau Benoît Barbotin fut l'inventeur d'une couronne d'engrenage destinée à manœuvrer la chaîne d'ancre, le barbotin. Pour l'anecdote, on notera qu'il était l'un des rescapés de la *Méduse*.

[5] «Les chaînes sont formées de maillons ovales (fig a) séparés en deux parties par un étai en fonte interposé d'une une [sic] branche à l'autre dans la direction du petit axe. Les maillons sont d'une longueur constante avec une tolérance de $\frac{1}{40}$ en plus ou en moins [...].

Les chaînes sont exécutées avec du fer en barre d'une qualité supérieure, elles sont composées de bouts de chaînes de 30 mètres de longueur qui sont réunis par un maillon d'assemblage avec boulon.» (Aubry 1866-1867 : 117-118)

[6] «Les chaînes se composent de trois pièces différentes : 1° Le chaînon, ou maillon, ou maille; 2° la manille ou pièce d'assemblage; 3° enfin l'émérillon [sic] ou pièce tournante.

La réunion d'un grand nombre de maillons forme la chaîne proprement dite, la manille et l'émérillon n'ayant d'utilité obligatoire que dans les câbles-chaînes, en usage dans la marine.» (Barbotin 1835 : 396)

[7] «Du chaînon, ou maillon, ou maille.

Il existe dans les câbles-chaînes trois espèces de maillons : 1° le maillon ordinaire, 2° le maillon sans étai, qui reçoit le collet de la manille, 3° le maillon à renfort, que je nomme aussi *maillon extrême*, dans lequel passe le boulon de la manille.» (Barbotin 1835 : 397)

[8] «Les câbles-chaînes se composent donc d'un certain nombre de morceaux de 30 mètres que l'on nomme aussi chaînons, réunis par des manilles ou pièces d'assemblage.

Ces pièces ont, comme on le sait, une forme différente de celle du maillon, et, par là, dérangent l'uniformité des chaînes.» (Barbotin 1835 : 399)

2.2 Stabilisation de l'usage

La première trace de la tendance des usagers à opter pour la monosémie s'observe, à notre connaissance, dans le *Manuel du gabier* (1885), où *maille* est présenté comme désignant «plus généralement» l'élément constitutif [9] et *maillon*, la portion de 30 mètres [10]. Cette distinction terminologique est scrupuleusement suivie à travers tout l'ouvrage [11].

[9] «479. *Nomenclature d'une chaîne*. — Une chaîne se compose de pièces de trois espèces différentes : 1° le chaînon, maillon ou plus généralement maille; 2° la manille; 3° l'émérillon [sic].

480. *Chaînon, chaîne ou maille*. — On distingue trois espèces de mailles : la maille à étai, la maille sans étai et la maille à renfort.» (*Manuel du gabier* 1885 : 343)

[10] «486. *Division d'une chaîne. Quelle est la longueur d'un maillon ?* — Les chaînes se divisent en bouts de 30 mètres ou 18 brasses, que l'on nomme *maillons*.

487. *Comment le maillon se termine-t-il à ses deux extrémités ?* — Le maillon se termine, à un bout, par une maille sans étai, à l'autre bout par une maille à renfort.» (*Manuel du gabier* 1885 : 346)

[11] «491. *Comment les chaînes sont-elles marquées?* — On marque les chaînes par maillons, de manière à connaître toujours la quantité de chaîne filée.

Le premier maillon ne se marque pas. Le second est marqué, sur la première maille à étau qui touche la maille à renfort, par plusieurs tours de fils de laiton passés autour de l'étau. Le troisième maillon est marqué de la même manière, sur l'étau de la deuxième maille à étau, et ainsi de suite.» (*Manuel du gabier* 1885 : 348)

Tous les manuels parus par la suite adopteront rigoureusement cette terminologie et leurs auteurs renonceront dès le début du XX^e s. à encore évoquer une éventuelle synonymie entre *maille*, *maillon* et *chaînon*, ce dernier terme quittant dès lors définitivement la sphère du vocabulaire nautique pour le sous-domaine concerné.

[12] «343. **Description d'une chaîne.** — Une chaîne ne forme point une série continue et indivisible de chaînons; elle peut se sectionner au contraire en bouts de 30 mètres ou 18 brasses nommés *maillons*.

Le maillon se compose principalement de chaînons ou mailles dites *mailles à étau*. La *maille à étau* est ainsi nommée parce qu'elle porte, en son milieu, une entretoise ou étau, qui a pour but d'empêcher la maille de s'aplatir, et d'éviter les coques dans les chaînes.» (Deloncle *et al.* 1891 : I-277)

[13] «382. **Description d'une chaîne.** — Une chaîne se subdivise en bouts de 30 mètres qui portent le nom de *maillons*.

Chaque maillon est composé de *mailles à étau*.» (Massenet *et al.* 1911 : 289)

[14] «Q. Quelle est la longueur d'un maillon ?

R. Trente mètres.

Q. Décrivez une chaîne d'ancre ?

R. La chaîne est étaliquée sur l'organeau de l'ancre au moyen d'une manille, puis vient une pièce de fer de forme spéciale, quelques mailles à étau, un émerillon [*sic*], quelques mailles à étau, une maille sans étau avec manille pour compléter le maillon, une manille à renfort commençant le premier maillon, puis une série de mailles à étau, séparées tous les trente mètres par une manille avec maille de renfort, jusqu'à l'étaliquure de cale ou le dernier maillon est sans étau.» (Coûteaux 1912 : 54-55)

[15] «166. **Chaînes.** — Une chaîne se divise en bouts de 30 mètres appelés *maillons*. Chaque maillon est composé de mailles à étau.» (*Manuel du gabier* 1917 : 155)

Les premières attestations d'un usage terminologique commun se trouvent donc dans des ouvrages publiés par l'autorité militaire et destinés à l'instruction des matelots ou des officiers (*Manuel du gabier* 1885 et Deloncle *et al.* 1891) avant d'apparaître dans des ouvrages destinés à l'instruction des officiers de la marine marchande (Massenet *et al.* 1911 et Coûteaux 1912)⁴.

4 Le système de l'inscription maritime alors en vigueur en France depuis Colbert fait de chaque marin ou marin-pêcheur une recrue de la flotte de guerre et implique un passage de plusieurs années par l'armée. À défaut de disposer d'études sérieuses sur ce sujet, on peut toutefois émettre l'hypothèse que la terminologie en usage à bord des navires de guerre a nécessairement dû influencer la langue des marins du commerce et de la pêche, voire constituer un mode récurrent d'implantation.

La distinction terminologique entre *maille* et *maillon* se retrouvera ensuite dans les textes des sociétés de classification (p.ex. le Bureau Veritas) et finira par atteindre un statut de norme dans les textes de l'ISO (ISO 3828 1984 et ISO 1704 1991) et de l'Organisation maritime internationale (OMI 1985). Bien entendu, cela ne signifie nullement que ces normes sont toujours parfaitement respectées dans l'usage quotidien, voire dans certains documents de grande diffusion⁵.

[16] 3.18	shackle of chain cable	27.5 m of	maillon	Longueur de chaîne
	shot of chain cable*	chain cable	de	correspondant à
	length of chain cable*		chaîne	27,5 m
				ISO 3828 (1984 : 3.18)

[17] «Toutes les dimensions, basées sur les diamètres nominaux des mailles ordinaires, s'entendent après soumission de la chaîne et des manilles aux charges d'essai légales.» (ISO 1704 1991 : 1)

3. L'USAGE DÉCRIT DANS LES DICTIONNAIRES SPÉCIALISÉS

3.1 Au XIX^e siècle

Les dictionnaires de marine du XIX^e siècle accusent, bien entendu, un retard par rapport aux progrès techniques. Si l'on retrouve déjà une mention du câble-chaîne (*s.v. câble*) dans le *Dictionnaire pittoresque de marine* de Jules Lecomte (1835 : 85), cité plus haut [1], et dans le célèbre *Glossaire nautique* d'Augustin Jal (1848 : I-375), ce n'est que dans le *Dictionnaire de marine à voile et à vapeur* de Bonnefoux & Pâris (1856) que chaque terme reçoit une entrée particulière. Les informations fournies constituent une radiographie exacte de l'usage observable dans les manuels parus antérieurement ou à la même époque et dont rendent compte les extraits [4] à [8].

[18] «MAILLE, s.f. [...] Maille est enfin syn. de Maillon (de câble-chaîne). P. l'exp. Voy MAILLON. P. U.» (Bonnefoux & Pâris 1856 : 483)

[19] «MAILLON, s.f. *Rolling hitch*. Sorte de nœud coulant ouvert en forme de cercle plus ou moins grand, que l'on fait avec un moyen ou petit cordage, pour relever un objet qui est au fond de l'eau et qui offre quelque prise. [...]

Les anneaux des câbles-chaînes s'appellent, d'ailleurs, des Maillons et, même, des Mailles; mais le premier de ces deux termes est le plus usité: il y a : 1° le Maillon ordinaire qui est, simplement, une sorte de boucle, et qui est destiné à recevoir le collet d'une manille; 2° le Maillon à renfort ou extrême, dans lequel passe le boulon d'une manille; 3° le Maillon à étau ou garni d'un étau : p.p.a.d. voy. au mot ÉTAI; 4°

⁵ Ainsi, les annexes du cours de Renier et Verstraeten (1925) contiennent un modèle non daté de «certificat d'épreuves de chaîne - certificate of chain» du Bureau Veritas dans lequel *maillon* est utilisé comme équivalent de *link* et *bout* comme équivalent de *length*. Ces usages sont en contradiction totale avec le dictionnaire *De la quille à la pomme de mâ*t (Paasch *et al.* 1937) auquel collabora le Bureau Veritas et font immédiatement songer à une erreur de traduction au départ de l'anglais. Plus récemment, le TC8/SC7 a utilisé *maillon* au sens de 'maille' dans la norme ISO 21 (1985), un an après la publication par le TC8/SC4 de la norme ISO 3828 (1984), consacrée au vocabulaire de la construction navale... On peut penser que ladite normes ISO 21 a été traduite en français à l'aide de l'un des dictionnaires évoqués en 3.3.

le Maillon dit d'Affourche ou tournant, ou Maillon d'Affourchage. P.p.a.d. voy. au mot ÉMÉRILLON.» (Bonnefoux & Pâris 1856 : 483)

[20] «CHAÎNON, s.m. Anneau d'un câble-chaîne, mais plus ordinairement appelé Maillon.

On nomme encore Chaînon la portion d'un câble-chaîne comprise entre deux manilles, et qui a une longueur d'environ 30 mètres.» (Bonnefoux & Pâris 1856 : 175)

[21] «CHAÎNE, s.f. [...] Le mot Chaîne tend à se substituer à celui de CÂBLE-CHAÎNE (P. l'exp. Voy. ce mot)» (Bonnefoux & Pâris 1856 : 175).

3.2 La première moitié du XX^e siècle

Le dictionnaire de Bonnefoux & Pâris peut être considéré comme le dernier grand dictionnaire de marine de la langue française avant la révision du très sérieux *Dictionnaire de l'océan* du Conseil international de la langue française (CILF 1989), lequel est toutefois davantage orienté vers l'hydrographie et l'océanologie⁶. Les seules exceptions notables sont deux ouvrages de dimensions plus modestes : le *Vocabulaire des termes de marine* de Soé *et al.* (1910) et le *Dictionnaire de marine* de Gruss (1978), sur lequel nous devons bientôt revenir.

Au XX^e siècle, les dictionnaires de marine sérieux seront généralement des dictionnaires polyglottes basés sur le modèle du novateur *De la quille à la pomme de mât* du capitaine Heinrich Paasch, édité par cinq fois entre 1885 et 1937 et qui occupera toute l'avant-scène jusqu'à la Deuxième Guerre mondiale (Van Campenhoudt 1994).

Germanophone de naissance, Paasch s'est parfois inspiré du dictionnaire de Bonnefoux & Pâris pour décrire la terminologie française. Ceci explique que dans les éditions du dictionnaire parues de son vivant (1885, 1894 et 1901), on trouve une description d'un usage déjà décrit par Bonnefoux & Pâris (1856 : *ibid.*), mais qui devient progressivement obsolète au regard des manuels de l'époque (*cf.* extraits [12] à [15]).

[22] Link of a Chain-cable. See Pl. 73 A. ⁴ .	Maillon de câble-chaîne. Voy. Pl. 73 A. ⁴ .	Ankerkettenglied; Ankerkettenschake. Siehe Pl. 73 A. ⁴ . (Paasch 1901 : 249)
--	--	--

[23] Length of a Chain-cable. Is the term given to a portion of chain-cable between two joining-shackles, usually 15 fathoms.	Chaînon. Terme donné à la partie d'un câble-chaîne, comprise entre deux manilles ayant ordinairement une longueur de 15 brasses.	Kettenlänge; Ein Schäkel Kette. Benennungen für den sich zwischen zwei Schäkeln befindenden Theil einer Ankerkette; gewöhnlich 15 Faden. (Paasch 1901 : 249)
--	---	---

⁶ L'honnêteté nous conduit à mentionner également la lente mais passionnante révision du *Glossaire nautique* de Jal par le C.N.R.S.

L'œuvre de Paasch fut perpétuée par une équipe internationale, mais il faudra attendre l'ultime remaniement de 1937, auquel collabora le célèbre Bureau Veritas, pour y lire une description qui corresponde à l'usage observé depuis la fin du XIX^e s. dans les manuels spécialisés.

[24] Maille de chaîne. Voy. Pl. 60. ⁴ & ⁶ .	Link of a chain-cable. See Pl. 60. ⁴ , ⁶ .	Ankerketten glied; Ankerketten schake. Siehe Pl. 60. ⁴ , ⁶ .	Eslabón de cadena. V. Lam. 60. ⁴ y ⁶ .	Maglia di catena d'ancora. Vedi Tav. 60. ⁴ & ⁶ .
--	---	---	---	---

(Paasch *et al.* 1937 : 189)

[25] Maillon. Partie d'une chaîne comprise entre deux manilles et ayant une longueur de 30 mètres environ.	Length of chain-cable. Is the term given to a portion of chain-cable between two joining-shackles, usually 15 fathoms.	Kettenlänge; Eine Länge Kette von Schäkel zu Schäkel; Benennungen für den sich zwischen zwei Schäkeln Befindenden Teil einer Ankerkette; gewöhnlich 15 Faden.	Grillete; Malla. Partes de una cadena comprendidas entre dos grilletes de unión, que tienen ordinariamente longitudes de 25 m. á 30 m.	Lunghezza di catena; Nodo di catena d'ancora. Parte di catena compresa tra due maniglie. Ha una lunghezza di circa 30 metri ed in Inghilterra di metri 27,43 (15 braccia).
---	---	--	---	---

(Paasch *et al.* 1937 : 189)

De nombreux dictionnaires de marine polyglottes publiés au XX^e siècle sont inspirés des principes macrostructurels adoptés par Heinrich Paasch : approche conceptuelle, classement logique, disposition par colonnes, recours systématique aux illustrations, etc. La plupart de ces ouvrages conçus dans un cadre professionnel font écho à l'usage contemporain en distinguant clairement les acceptions de *maille* et de *maillon* (IMCO 1963; Segditas 1965-1966; A.I.P.C.N. 1966; Verhaege 1994)⁷

3.3 Après 1940 : propagation d'une confusion

Lorsqu'on consulte les nombreux dictionnaires de marine parus après 1940, on est étonné de découvrir que certains d'entre eux décrivent à nouveau *maille* et *maillon* comme des synonymes. D'autres ouvrages, plus nombreux, conservent toutefois la distinction terminologique rigoureuse qui s'était progressivement mise en place au fil du XIX^e siècle.

Le tableau ci-dessous donne une idée de l'ampleur du phénomène pour le concept 'anneau d'une chaîne'.

⁷ Par contre, Liévin Groenen (1939 : 286) reprend l'information surannée proposée dans les quatre premières éditions du dictionnaire de Paasch. Il est vrai que son *Encyclopédie illustrée de la marine* est largement fondée sur le pillage du travail de son illustre prédécesseur.

<i>maille</i> Δ <i>maillon</i>	
Gruss (1945, 1952 & 1978)	Gruss (1943)
Hazard (1951)	Kerchove (1961)
Merrien (1962)	IMCO (1963)
Vandenberghé et Chaballe (1978)	Segditas (1965-1966)
Commission of the European Communities (1992a)	A.I.P.C.N. (1966)
Vandenberghé et Johnen (1994)	Sizaire (1978)
B.T.Q. (1995)	Ratcliff (1983)
EURODICAUTOM (1997)	ISO 3828 (1984)
	O.M.I. (1985)
	CILF (1989)
	Dobenik & Hartline (1989)
	Bruno & Mouilleron-Becar (1991 & 1994)
	Verhaege, L. (1994)

Parmi tous ces ouvrages, le *Dictionnaire de l'océan* du CILF (1989 : 291) est l'un des rares à signaler l'existence d'une confusion entre les deux termes [26].

[26] «**MAILLE** n.f.

En. **Link of a chain**

De. **Schäkel**

Es. **Eslinga**

Techn.

Anneau d'une chaîne d'ancre.

Il existe des mailles sans étai, ou plates, et des mailles avec étai, c'est-à-dire renforcées dans la partie centrale.

Syn. Maillon (impropre).» (CILF 1989 : 291)

La mention «impropre» ne permet pas de savoir dans quels contextes cette synonymie est observable. Pour ce qui est des dictionnaires, la chose semble évidente au vu du tableau qui précède. À l'écrit, par contre, la distinction entre *maille* et *maillon* se maintient bel et bien dans les manuels de l'après-guerre et est même confirmée — nous l'avons vu en 2.2 — par des normes. Nous n'avons observé une confusion des deux termes que dans des textes propres au domaine, fort spécifique, de la plaisance, où la longueur des chaînes d'ancre dépasse rarement 30 mètres et où la distinction entre les deux concepts n'est donc pas pertinente.

Pour ce qui concerne l'usage oral, notre enquête auprès de quelques officiers de la marine marchande — jeunes ou vieux, belges ou français — atteste une bonne connaissance de la distinction entre *maille* et *maillon*. Comme l'ont fait observer plusieurs témoins belges francophones, la proximité des termes *maille* et *maillon* ne met toutefois personne à l'abri d'une confusion... légitimée, il est vrai, par certains dictionnaires. Par ailleurs, la réalité des équipages actuels — largement recrutés dans le tiers monde —

conduit à un grand mélange de langues à bord des navires et à la prédominance d'un anglais qui tient plus du sabir que de la noble langue de Masefield et de Conrad. Ainsi, les marins belges interrogés utilisent volontiers l'anglais *shackle* pour désigner la longueur de chaîne ou l'élément réunissant deux longueurs de chaîne⁸.

De ce point de vue, il convient de souligner que notre étude porte sur des pratiques terminographiques et non sur ce que doit ou devrait être la terminologie nautique de langue française, parlée par un nombre toujours plus restreint de marins et de pêcheurs.

4. ORIGINE ET DIFFUSION D'UNE CONFUSION

La comparaison systématique d'un grand nombre de dictionnaires de marine publiés depuis 1830 nous a convaincu que la synonymie entre *maille* et *maillon* apparue dans certains dictionnaires après la dernière guerre n'était pas la trace de l'émergence d'un nouvel usage, d'ailleurs nullement confirmé par la citation de contextes ou par des marques d'emploi. Elle est, en fait, le simple résultat d'un travail de copiste...

4.1 Première attestation

L'observation du tableau comparatif produit ci-dessus montre que le premier dictionnaire spécialisé à présenter à nouveau *maille* et *maillon* comme synonymes est la deuxième édition du *Petit dictionnaire de marine* de Robert Gruss (1945). Paradoxalement, la première édition (Gruss, 1943) n'adoptait pas le même point de vue [27-28].

[27] «**Maille.** — Intervalle qui se trouve entre deux couples voisins d'un navire ou entre deux varangues.

Ouverture laissée entre les fils des filets de pêche.

Anneau d'une chaîne.» (Gruss 1943 : 129)

[28] «**Maillon** — Partie d'une chaîne comprise entre deux manilles et ayant une longueur de 30 mètres environ.

Un navire qui mouille une ancre file d'autant plus de maillons de chaîne qu'il veut s'assurer une bonne tenue en cas de mauvais temps.» (Gruss 1943 : 129)

Dans la seconde édition, le contenu de l'article *maillon* [29], enrichi des acceptions 'anneau d'une chaîne' et 'nœud coulant' (cf. note 1), est en réalité immédiatement inspiré de la définition fournie sous une même entrée par Bonnefoux & Pâris (1856 : 483) et déjà citée plus haut [19]. Comme on le verra en 4.3, Gruss a reconnu s'être inspiré du dictionnaire de ces auteurs ainsi que de celui de Paasch⁹.

8 Les chaînes ont d'abord été adoptées au Royaume-Uni, pays d'où elles étaient importées avant que la France n'arrive à en produire. En anglais, le même terme *shackle* désigne aussi bien la pièce (*manille*) réunissant les longueurs de chaîne que la longueur de chaîne proprement dite (*maillon*). Il s'agit d'une extension de sens profondément logique, puisque l'on compte la longueur de chaîne filée en fonction du nombre de manilles que l'on a vu sortir du puits aux chaînes.

9 Le copiage des définitions apparaît dès la première édition. Ainsi, la définition de *chaîne* (Gruss, 1943 : 48) équivaut quasiment mot pour mot à celle proposée par Paasch (1901 : 248). Parmi les entrées voisines de *maillon* dans le dictionnaire de Gruss (1945), on trouve p.ex. des emprunts à Bonnefoux & Pâris (1856) pour *macaron*, *magasin général*, *mailler*, etc.

[29] «**Maillon** — 1° Partie d'une chaîne comprise entre deux manilles et ayant une longueur de 30 mètres environ.

Un navire qui mouille une ancre file d'autant plus de maillons de chaîne qu'il veut s'assurer une bonne tenue en cas de mauvais temps.

2° Anneau d'une chaîne, chaînon.

3° Nœud coulant que l'on fait avec un moyen ou petit cordage, pour relever un objet qui est au fond de l'eau et qui offre quelque prise.» (Gruss 1945 : 125)

La consultation des ouvrages spécialisés rédigés en français après la Deuxième Guerre mondiale ne confirme pas l'émergence d'une nouvelle confusion de *maille* et de *maillon*, dont Gruss aurait voulu rendre compte. Ainsi, le manuel de construction navale de Chicot (1960) distingue-t-il clairement les deux termes :

[30] «Les chaînes d'ancre sont composées de mailles soudées avec étai en acier. Le calibre (D) de la chaîne est le diamètre du fer rond qui constitue la maille. Le pas est la longueur intérieure de la maille, égal à 4 D pour les chaînes de type international.

[...]

Les chaînes sont fabriquées par tronçons de 30 mètres environ de longueur, appelés *maillons*. Les maillons sont reliés entre eux par des manilles de jonction ou des mailles démontables (mailles Kenter).» (Chicot 1960 : 291)

Tout semble donc indiquer que c'est bien la description d'un usage ancien, réalisée 89 années auparavant par Bonnefoux & Pâris (1856), qui inspire Gruss (1945) lorsqu'il modifie la définition de *maillon* dans la deuxième édition de son *Dictionnaire de marine*. Cette même modification demeurera telle quelle dans la dernière édition de 1978 [31], rédigée 122 ans après la parution du dictionnaire de Bonnefoux & Pâris¹⁰ !

[31] «**Maillon**, *m. Schackle*. 1°) Partie d'une chaîne d'ancre comprise entre deux manilles et d'une longueur de 30 mètres. 2°) Anneau d'une chaîne, chaînon. 3°) Nœud coulant fait avec un petit cordage pour relever un objet immergé offrant quelque prise.» (Gruss 1978 : 191)

4.2 Dissémination de la confusion

L'analyse des autres dictionnaires qui établissent une synonymie entre *maille* et *maillon*, donne à penser que c'est bien le crédit attaché au travail de Gruss qui est à l'origine de la diffusion de cette confusion.

Le fait est flagrant dans le cas des grandes bases de données terminologiques que sont EURODICAUTOM et la Banque de terminologie du Québec¹¹. Il faut rappeler que

10 Gruss a pu également être conforté dans sa description par la lecture du *Nouveau Larousse illustré* (1897-1904) ou de son «quasi-clone» pour le domaine nautique, le *Larousse du XX^e siècle* (1928-1933). En effet, pour les termes concernés, ces grandes encyclopédies accusent, elles aussi, un important retard par rapport à l'usage.

11 *Termium* (1996) échappe à la critique puisque le cédérom ne propose qu'une fiche consacrée à *maillon* 'portion de chaîne de 15 brasses' (fiche n° 508793). Nous n'avons pas consulté le cédérom *Termdok2* (1992) dans la mesure où les données sur la langue française y sont généralement extraites de *Termium* et d'EURODICAUTOM.

ces bases ont été essentiellement constituées par l'accumulation de données tirées de dictionnaires déjà publiés, ce qui implique la présence d'une information parfois redondante, mais aussi parfois contradictoire.

4.2.1 EURODICAUTOM

Pour les concepts qui nous occupent, EURODICAUTOM propose des fiches issues du *Dictionnaire technique illustré* de l'Association internationale permanente des congrès de navigation (A.I.P.C.N. 1966 : IV-74) [32-33] et de l'imposant *Multilingual dictionary of fishing vessels and safety on board* [34-35] publié par la Commission des Communautés européennes en collaboration avec de nombreux experts (Commission of the European Communities 1992a : n^{OS} 2586 et 2819).

Le risque évident est que, du fait de l'enrichissement de la base de données au départ de différents dictionnaires, la divergence de la terminologie en fonction des sous-domaines ne semble pertinente : à la lecture des fiches reproduites ci-dessous¹², *maillon* apparaît comme synonyme de *maille* 'anneau d'une chaîne' dans le secteur des pêches, mais non dans celui du transport maritime ou fluvial. Vu la nature des sources, qui ne sont en rien spécifiques à la pêche, sinon qu'elles sont cautionnées par la présence d'experts de ce domaine, on peut sérieusement douter du fondement d'une telle distinction : le dictionnaire de Gruss traite de la marine en général, de même que la norme ISO 3828 (1984) également citée. Bien entendu, ceci soulève de graves questions quant à la pertinence des sources dictionnairiques et à la nature du travail de l'expert...

[32]

Reliability code : 3

Date : 940803

ID Number : 0001131

Type : Dict. techn. illustré (AIPCN) : IV-Bateaux (=CNB69)

Terminology Bureau : Bureau de terminologie de Luxembourg (Terminology Office, Luxembourg, European Commission), SdT/AGL03-Lx (=BTL)

Subject code : materials science (=TE6) seaborne traffic (=TR7)

French

Keyword : maille

Reference : Dict. Techn. ill. IV-74

English

Keyword : link

Reference : Dict. Techn. ill. IV-74

[...]

[33]

Reliability code : 3

Date : 940803

ID Number : 0001132

Type : Dict. techn. illustré (AIPCN) : IV-Bateaux (=CNB69)

Terminology Bureau : Bureau de terminologie de Luxembourg (Terminology Office, Luxembourg, European Commission), SdT/AGL03-Lx (=BTL)

Subject code : inland waterway traffic (=TR6) seaborne traffic (=TR7)

12 Par souci d'économie, nous ne citons que les champs consacrés au français et à l'anglais.

French

Keyword : 1) maillon; 2) maillon de chaîne

Reference : 1) Dict.Techn.ill.IV-74; 2) OMI, Gloss Termes techniques, Londres 1982

English

Keyword : 1) cable length; 2) chain cable length

Reference : 1) Dict.Techn.ill.IV-74; 2) IMO, Gloss Termes techniques, London 1982

[...]

[34]

Reliability code : 4

Date : 950308

ID Number : 0002586

Type : Gloss. Navires de pêche (BTL) (=TOR93)

Terminology Bureau : Bureau de terminologie de Luxembourg (Terminology Office, Luxembourg, European Commission), SdT/AGL03-Lx (=BTL)

Subject code : fishing, fish breeding & aquaculture (=AGB)

French

Keyword : maillon; maille

Definition : Anneau allongé d'une chaîne.

Reference : Gruss, Dict. de Marine

English

Keyword : link

Definition : One ring or loop of chain.

Reference : A. Morrall, BMT

[...]

[35]

Reliability code : 4

Date : 950308

ID Number : 0002819

TypeGloss : Navires de pêche (BTL) (=TOR93)

Terminology Bureau : Bureau de terminologie de Luxembourg (Terminology Office, Luxembourg, European Commission), SdT/AGL03-Lx (=BTL)

Subject code : fishing, fish breeding & aquaculture (=AGB)

French

Keyword : maillon de chaîne

Definition : Longueur de chaîne correspondant à 27,5 m.

Reference : ISO/DIS 3828

English

Keyword : 1) shot of chain cable; 2) length of chain cable*; shackle of chain cable*

Definition : 27,5 m length of chain cable.

Reference : ISO/DIS 3828

Explanation : *) deprecated terms

[...]

4.2.2 La banque de terminologie du Québec

Lorsqu'on consulte les fiches de diffusion du cédérom *Le Doc*, de la B.T.Q. (1995), on découvre une information d'excellente qualité dans deux fiches [36-37] fondées sur la

synthèse de différentes sources et accompagnées d'excellentes notes précisant les risques de confusion¹³. Dans le même temps, une fiche particulière du fichier de diffusion [38] est consacrée à une information contenue dans le seul dictionnaire de Gruss (1978). Cette même référence justifiait déjà la synonymie proposée entre *maille* et *maillon* dans EURODICAUTOM (1997 : n° 0002586) et dans le glossaire des pêches de la Commission (Commission of the European Communities 1992a : n° 2586).

Dans le cas de cette dernière fiche, on notera la grave erreur de désignation du concept 'anneau d'une chaîne' en anglais. L'équivalent adéquat est *link* et non point *shackle*. *Shackle* désigne en anglais soit la manille 'anneau de liaison', soit le maillon 'longueur de 15 brasses'. L'erreur provient de la copie aveugle de Gruss, qui sous l'entrée *maillon* [31], ne propose que l'équivalent *shackle* pour désigner trois concepts différents...

[36]

Domaine générique : marine

spécifique : matériel de mouillage

ANGLAIS : link 1010157 0096461

Note : [Dans la source citée, l'entrée se lit comme suit :] link of a chain-cable. 10096461

FRANÇAIS : maille 1010157 011697 010676 0096461

Informations : n f

Définition : Anneau d'une chaîne d'ancre muni d'un étai pour éviter sa déformation. 10116971

[...] anneau constitutif d'une chaîne. La force, le calibre d'une chaîne se mesurent à l'échantillonnage de la maille, c'est-à-dire au diamètre et à la qualité d'acier employé. 10106761

Note : [Dans la source citée, l'entrée se lit comme suit :] maille et maillon. Il faut éviter d'employer comme synonymes les mots maille et maillon. Une fausse-maille est au contraire, une maille d'assemblage permettant de réunir deux mailles ou deux sections de chaîne. 10106761

Rédacteur : Office de la langue française

Date : 1984

Fichier : diffusion

[37]

Domaine générique : marine

spécifique : matériel de mouillage

ANGLAIS : shackle 1009646 000922 0057701

Définition : A length of cable or anchor chain ususally 15 feet. 10057701

FRANÇAIS : maillon 1009646 0116971

Informations : n m

Définition : Partie d'une chaîne d'ancre comprise entre deux manilles et d'une longueur de 30 mètre. 10096461

Élément d'une chaîne d'ancre d'une longueur égale à 30 mètre quel que soit la calibre des mailles. 10116971

[...] longueur de chaîne d'un seul lot. Cette longueur est de 30 m quelque soit l'échantillon de la chaîne. 10106761

13 Les nombreuses coquilles et fautes d'orthographe observables dans les extraits cités sont déjà présentes dans les fiches du cédérom.

Note : [Dans la source citée, l'entrée se lit comme suit :] maille et maillon. Il faut éviter d'employer comme synonyme les mots maille et maillon. Un faux-maillon est une section de chaîne de longueur inférieure à 30 m. |010676|
Synonyme 1 : maillon de chaîne |000922|; n m; Définition : La chaîne au bout de laquelle est l'ancre se compose de fraction d'environ 30 m. Chaque fraction est un maillon de chaîne. |000922|; NOTE : Ce mot peut évidemment conduire à méprise si l'on considère qu'en fait le maillon est composé de... mailles. Un navire a généralement 8 à 10 maillons de chaque bord dans son «puits» aux chaînes. |000922|
Rédacteur : Office de la langue française
Date : 1984
Fichier : diffusion

[38]

Domaine générique : marine
spécifique : matériel de mouillage
ANGLAIS : shackle |009646|
FRANÇAIS : maillon |009646|
Informations : n m
Définition : Anneau d'une chaîne, chaînon. |009646|
Rédacteur : Gruss, Robert
Date : 1978
Fichier : diffusion

4.3 Du crédit accordé au dictionnaire de Gruss

L'utilisation intensive du *Petit dictionnaire de marine* de Gruss comme source de référence privilégiée aura de quoi étonner. Il est vrai que, comme nous l'avons signalé en 3.2, ce dictionnaire de marine est l'un des rares à avoir été rédigés en français¹⁴ au cours de ce siècle et, à notre connaissance, le seul à couvrir de nombreux domaines de l'art de la navigation. Publié pour la dernière fois en 1978, il a connu un réel succès commercial. Il est toutefois malaisé de se prononcer sur les fondements du crédit accordé à cet ouvrage, notamment par des membres de l'Académie de marine (Mounier et Sizaire 1979).

À défaut d'avoir pu exercer la carrière maritime à laquelle il aspirait, Robert Gruss consacra toute sa vie active aux Éditions maritimes et d'outre-mer¹⁵. Passionné de plongée et de navigation, il fut l'auteur de divers ouvrages, dont le célèbre *Petit dictionnaire de marine*. Dans l'introduction de la dernière édition [39], l'auteur précise bien quelles furent ses sources les plus importantes :

[39] «L'auteur d'un dictionnaire, général ou spécialisé, même rédigé sans prétention tel que celui-ci, ne peut l'entreprendre sans s'appuyer sur les travaux de ses devanciers. Pour la période moderne (XIX^e et XX^e siècles) deux ouvrages essentiels sont à la base du présent travail : le *Bonnefoux* et le *Paasch*.» (Gruss 1978 : 7)

Le fait même que Gruss (1978 : 7) affirme dans sa préface s'être appuyé sur ces deux œuvres majeures publiées l'une en 1856 (Bonnefoux & Paris), l'autre entre 1885 et

14 Plusieurs dictionnaires rédigés en langue étrangère ont été ensuite traduits en français.

15 Ancienne Société d'éditions géographiques, maritimes et coloniales, elles-mêmes anciennes éditions Challamel.

1937 (Paasch) contribue sans doute à rendre son travail crédible aux yeux d'autres terminographes. Dans le même temps, ceux-ci préfèrent vraisemblablement citer un dictionnaire récent tel celui de Gruss, plutôt que les ouvrages anciens dont il est largement inspiré...

Si l'auteur fait également référence aux nombreux experts qui l'ont aidé, parmi lesquels on ne compte aucun linguiste, il ne présente jamais son dictionnaire comme une stricte radiographie de l'usage en synchronie. Son souci de s'appuyer sur des dictionnaires nés au siècle passé atteste une volonté de rendre compte d'emplois plus anciens, même si ceux-ci ne sont ni datés, ni présentés comme tels. Il est vrai que, comme de nombreux autres auteurs de dictionnaires spécialisés, Gruss est un amateur éclairé qui ne dispose d'aucune formation en lexicologie ou en terminologie. Ces remarques semblent fort importantes face au crédit accordé par de nombreux terminographes à un dictionnaire «dont le but et l'ambition ont toujours été de satisfaire le lecteur d'un ouvrage de marine ou d'un article de journal» (Gruss 1978 : 7).

Il convient, *in fine*, d'observer que tous les auteurs de dictionnaires n'ont pas copié Gruss. On remarquera avec intérêt que la majorité des dictionnaires de marine de l'après-guerre qui distinguent correctement *maille* et *maillon* ont été rédigés par des professionnels, dont certains combinent adéquatement une formation de marin et de linguiste. Ce ne sont malheureusement pas toujours leurs ouvrages qui sont retenus dans les grandes bases de données terminologiques.

5. QUELQUES RÉFLEXIONS

À défaut de permettre de tirer des conclusions générales, cette étude restreinte et isolée ouvre des pistes de réflexion quant à la nature des dictionnaires spécialisés, à leur méthodologie, à leurs auteurs et collaborateurs, à leurs sources ou encore à leur influence sur l'usage.

5.1 Dictionnaires de marine et pratique terminographique

Au regard des critères méthodologiques de la terminologie, on pourrait aisément balayer d'un geste de la main nombre de dictionnaires de marine d'hier et d'aujourd'hui. Notre étude a, il est vrai, couvert tous les types d'ouvrages : dictionnaires monolingues adoptant plutôt une démarche encyclopédique, dictionnaires bilingues ou multilingues trop souvent restreints à des listes d'équivalents, bases de données terminologiques. Nous nous devons de faire référence à ces ouvrages dans leur diversité, car ils sont ceux-là mêmes auxquels les marins se réfèrent, avec plus ou moins de bonheur.

Le traducteur se tournera naturellement vers les dictionnaires bilingues ou multilingues. Selon son choix, il tombera sur des ouvrages excellents ou médiocres, la réputation de l'éditeur n'ayant malheureusement aucune influence sur la qualité de l'œuvre. Aujourd'hui aisément accessibles, sur CD-ROM ou via Internet, les grandes bases de données terminologiques sont assurément celles qui lui offriront le plus de renseignements de nature linguistique (références, notes...). Malheureusement, du fait de leur constitution par accumulation de données, et en dépit de leur nettoyage régulier, elles offrent un mélange du meilleur et du pire. De ce point de vue, la meilleure garantie de fiabilité de leur utilisation est encore le regard critique du traducteur — dès lors nécessairement spécialiste

du domaine — ou l'intervention de terminologues patentés qui prennent le temps de travailler dans les règles de l'art.

De ce point de vue, il convient de remettre en cause la pratique terminographique — parfois encore enseignée, voire présentée comme une exigence — qui consiste à mentionner des termes et des définitions attestés dans des dictionnaires. Au minimum, ces informations doivent être confirmées par des attestations. Force est, hélas, de constater qu'aucun des dictionnaires, aucune des bases de données consultés ne proposent le moindre contexte de manière systématique¹⁶. Les bases de connaissance devraient bientôt permettre de confronter le contenu de chaque fiche terminologique à l'usage écrit, observable dans de vastes corpus. Face à la puissance de tels outils d'avenir, il convient que les terminographes s'interrogent sur la pertinence d'une pratique, sinon d'une méthodologie, trop souvent fondée sur la citation des prédécesseurs.

Préalablement à la description terminographique d'un domaine, il convient aussi d'analyser les filiations entre les dictionnaires et donc d'adopter une perspective diachronique. Hélas, le terminologue est rarement formé à la critique des documents. La copie des devanciers et l'usage de la référence bibliographique comme argument d'autorité nuisent assurément à la reconnaissance de la terminologie au sein du champ de la linguistique. Dans l'enseignement universitaire, l'étude diachronique des vocabulaires spécialisés relève traditionnellement de la lexicologie et constitue une source inépuisable de sujets de thèses et de mémoires; sans doute n'envisage-t-on pas assez l'intérêt d'une telle démarche pour le terminologue, réputé œuvrer en synchronie... dans l'urgence perpétuelle.

5.2 Le rôle des experts

Les glossaires des pêches de la Commission des Communautés européennes et les fiches d'EURODICAUTOM qui en sont immédiatement extraites, ont été réalisés avec l'aide d'un expert du domaine pour chaque langue. La version publiée (Commission of the European Communities, 1992a) mentionne le nom de ces collaborateurs, mais ne reproduit pas les champs «référence», en sorte que le lecteur croit lire l'avis d'un expert et non le contenu d'un autre dictionnaire, dont nous ignorons d'ailleurs s'il a été soumis audit expert...

En fait, les professionnels d'une discipline sont eux-mêmes des grands consommateurs de dictionnaires, qu'ils compulsent lorsqu'ils désirent désigner un concept avec certitude dans le cadre d'un usage public (rapport, cours, conférence...). Tout terminologue qui a consulté plusieurs professionnels d'un même domaine, voire qui les a réunis, a pu plus d'une fois mesurer l'ampleur de leurs contradictions et de leurs propres interrogations.

Le professionnel d'un domaine n'est pas toujours, loin s'en faut, un expert de la terminologie de ce domaine. En termes de linguistique de terrain, il est d'abord et avant tout un informateur. L'interroger exige une méthodologie particulière, rarement enseignée

¹⁶ On notera au crédit de Gruss (1978) la présence de citations essentiellement littéraires dans quelques articles.

au terminologue. Connaître ses dictionnaires de référence, en étudier la valeur et la filiation est, de même, une tâche indispensable pour l'enquêteur.

5.3 L'influence du dictionnaire sur l'usage

La prise en compte de la diachronie sur une période d'un siècle et demi nous a permis d'assister, au départ d'une concurrence entre plusieurs synonymes, à l'émergence progressive d'une terminologie. Une fois largement implantée, celle-ci est finalement consacrée par la normalisation au sein d'organismes de référence (sociétés de classification, Organisation maritime internationale, ISO...). Ce processus correspond à celui déjà fort bien décrit par Éloy et Humbley (1993 : 14) :

[40] «Nous distinguons ici deux types de besoin qui se font ressentir lors de la création d'un vocabulaire nouveau : besoin lexical et besoin terminologique. Le premier est souvent comblé par une profusion synonymique de différents ordres, que nous examinerons; le second relève plutôt de l'épuration, de la décision plus ou moins consciente de ne retenir qu'un seul terme pour dénommer un concept reconnu dans la nouvelle sphère d'activités.»

Lorsqu'une technologie est fort récente, il convient de considérer avec prudence les tentatives d'aménagement et de normalisation. Dans le cas de la chaîne d'ancre, comme dans bien d'autres déjà étudiés, on s'aperçoit que la stabilisation de l'usage prend un certain temps et s'opère naturellement, au travers de facteurs particuliers, de vecteurs pertinents. Le rôle de l'enseignement militaire semble ici prédominant (*cf.* note 4) et rappelle la place de l'armée dans la mise en place du vocabulaire de l'aviation (Guilbert 1965 : 87, 196)¹⁷.

Le succès du dictionnaire de Gruss comme source de référence, même auprès des professionnels du domaine, pourrait, par effet de dissémination, remettre en cause ce processus d'implantation naturelle de la terminologie et sa validation finale comme norme. Lorsque l'erreur d'un terminographe connaît un tel succès dans les autres dictionnaires, on en vient d'ailleurs à se demander si elle ne s'en trouve pas légitimée. Il est toutefois trop tôt pour se prononcer sur la pénétration dans l'usage de la confusion entre les termes *maille* et *maillon*, notamment par le biais de la traduction (*cf.* note 5).

Dans l'histoire d'une langue, et même d'une langue spécialisée, certains contresens connaissent parfois un franc succès, au point de finir par s'imposer et de constituer une nouvelle norme. Il est, somme toute, rassurant d'observer qu'à l'heure de la terminotique, des erreurs de nature humaine puissent encore influencer le contenu des dictionnaires électroniques, voire se répandre dans l'usage du fait de la puissance de dissémination des outils nés des progrès de l'ingénierie linguistique.

17 On observera d'ailleurs que la distinction *maille* - *maillon* est toujours présente dans les dictionnaires rédigés dans un cadre militaire, tels ceux de Ratcliff (1983) et de Dobenik et Hartline (1989).

RÉFÉRENCES

- A.I.P.C.N. (1966) : *Dictionnaire technique illustré en six langues*. Vol. IV : *Bateaux et navires, propulsion, conditions de navigation, jauge, remorquage, touage, halage, bacs*, Bruxelles. Association internationale permanente des congrès de navigation.
- AUBRY, ? (1866-1867) : *Principes élémentaires sur lesquels repose la manœuvre des vaisseaux*, Brest. École navale impériale, Impr. Roger Père (polycopié de l'année scolaire 1866-1867).
- BARBOTIN, B. (1835) : «Première partie du rapport de M. Barbotin, capitaine de corvette, sur son système de cabestan à double engrenage», dans *Annales maritimes et coloniales — Partie non officielle*, to I, pp. 387-424.
- BONNEFOUX, P.-M.-J. de et E. PÂRIS (1856) : *Dictionnaire de marine à voile et à vapeur*, 2^e édit., Paris, Bertrand, 2 vol.
- BRUNO, A. et Cl. MOUILLERON-BECAR (1991) : *Dictionnaire maritime thématique anglais et français*, Paris, Milan, Barcelone et Bonn, Masson.
- BRUNO, A. et Cl. MOUILLERON-BECAR (1994) : *Dictionnaire maritime thématique anglais et français*, 2^e édit. revue et augmentée, Paris, Milan et Barcelone, Masson (Bibliothèque de l'Institut français d'aide à la formation professionnelle et maritime).
- B.T.Q. (1995) : *Le Doc de l'Office de la langue française. La banque de terminologie du Québec (B.T.Q.). La banque documentaire. Le français au bureau*, banque de données sur cédérom, Québec, Office de la langue française.
- BUSSY, Ch. (1862) : *Dictionnaire universel de marine avec traduction des termes de la marine française en termes de la marine anglaise*, Paris, Firmin Didot.
- CHICOT, E. (1960) : *Construction du navire de commerce*, Paris, Éditions maritimes et coloniales.
- CILF (1989) : *Dictionnaire de l'océan*, Paris, Conseil international de la langue française.
- Commission of the European Communities (1992a) : *Multilingual dictionary of fishing vessels and safety on board*, 2nd ed., Oxford, Fishing News Books et Luxembourg, Office for Official Publications of the European Communities.
- Commission of the European Communities (1992b) : *Multilingual dictionary of fishing vessels and safety on board*, 2nd ed., Oxford, Fishing News Books et Luxembourg, Office for Official Publications of the European Communities.
- COÛTEAUX, H. (1912) : *Petit manuel de manœuvres*, Bruxelles, Larcier.
- DELONCLE ?, LAJARTE ? de et V. de KERRADEL (1891) : *Manuel du manœuvrier à l'usage des élèves de l'École navale et de l'École d'application*, Paris, Challamel, 3 vol.
- DOBENIK, R. et G. HARTLINE (1989) : *Dictionnaire technique de la marine. Anglais/français et français/anglais*, Paris, La Maison du dictionnaire.

- ÉLOY, J.-M. et J. HUBLEY (1993) : «La notion de besoin terminologique et la naissance de lexiques spécialisés au 19^e et au 20^e siècles», dans *Terminologies nouvelles*, n° 9, p. 14-18.
- EURODICAUTOM (1997) : *Eurodicautom*, base de données sur Internet, Luxembourg, Commission des Communautés européennes, <http://www2.echo.lu/edic/>
- GOCVIC, E. et H.G. JANSEN (1844) : *Dictionnaire universel, historique et raisonné français-hollandais de marine et de l'art militaire*, La Haye et Amsterdam, Van Cleef.
- GROENEN, L. (1939) : *Encyclopédie illustrée de la marine. Anglais - français - néerlandais*, Anvers, J. Guillaume.
- GRUSS, R. (1943) : *Petit dictionnaire de marine*, Paris, Société d'éditions géographiques, maritimes et coloniales.
- GRUSS, R. (1945) : *Petit dictionnaire de marine (...)*, édit. entièrement refondue et augmentée, Paris, Société d'éditions géographiques, maritimes et coloniales.
- GRUSS, R. (1952) : *Petit dictionnaire de marine (...)*, 3^e édit. entièrement refondue et augmentée, Paris, Société d'éditions géographiques, maritimes et coloniales.
- GRUSS, R. (1978) : *Dictionnaire Gruss de marine*, 4^e édit., Paris, Éditions Maritimes et d'outre-mer.
- GUILBERT, L. (1965) : *La formation du vocabulaire de l'aviation*, Paris, Larousse.
- HAZARD, J. (1951) : *Dictionnaire anglais-français et français-anglais des termes de marine et du navire (pont et machine)*, 1^{re} édit., Paris, Éditions maritimes et coloniales.
- IMCO (1963) : *Glossary of maritime technical terms = Glossaire des termes techniques maritimes = Glosario de terminos tecnicos maritimos = Slovare norskih tehniceskih termino*, London, Inter-Governmental Maritime Consultative Organization.
- Instructions* (1833) : *Instructions sur le linguet de chaîne, sa manœuvre et son entretien*, Paris, Imprimerie Royale, 15 p.
- ISO 3828 (1984) : *Construction navale et structures maritimes — Auxiliaires de pont — Vocabulaire*, 2^e édit., Genève, Organisation internationale de normalisation (ISO, TC 8/SC4).
- ISO 1704 (1991) : *Construction navale — Chaînes d'ancre à mailles étauçonnées*, Genève, Organisation internationale de normalisation (ISO, TC 8/SC4).
- ISO 21 (1985) : *Construction navale — Navigation intérieure — Barbotins pour les chaînes d'ancre étauçonnées*, Genève, Organisation internationale de normalisation (ISO, TC 8/SC7).
- JAL, A. (1848) : *Glossaire nautique. Répertoire polyglotte de termes de marine anciens et modernes*, Paris, Didot, 2 vol.
- KERCHOVE, R. de (1961) : *International maritime dictionary. An encyclopedic dictionary of useful maritime terms and phrases, together with equivalents in French and German*, 2nd

edition, New York, Cincinnati, Toronto, London, Melbourne, Van Nostrand Reinhold Company.

LECOMTE, J. (1835) : *Dictionnaire pittoresque de marine*, Paris, Bureau central de la France maritime.

Manuel du gabier (1885) : *Manuel du gabier publié par ordre de M. le ministre de la marine et des colonies*, 3^e édition, Paris, Librairie militaire de L. Baudoin et cie.

Manuel du gabier (1917) : *Manuel du gabier*, 6^e édition, Paris, Imprimerie militaire.

Manuel du manœuvrier (1891) : *Manuel du manœuvrier à l'usage des élèves de l'École navale et de l'École d'application*, Paris, Challamel. 3 vol.

MASSENET, G., VALLEREY, J. et A. LETALLE (1911) : *Gréement, manœuvre et conduite du navire à voiles et à vapeur*, Paris, Augustin Challamel, 2 vol.

MONTFERRIER, A.-S. de (1841) : *Dictionnaire universel et raisonné de marine*, Paris, Bureau du dictionnaire de marine.

MOUNIER, P. et P. SIZAIRE (1979) : *Bref historique du dictionnaire de marine de ses origines à nos jours*, Paris, Académie de marine.

Nations unies (1992) : *Transports maritimes. Termes techniques, juridiques et commerciaux. Lexique anglais-français*, New York, Nations unies.

O.M.I. (1985) : *Vocabulaire normalisé de la navigation maritime*, Londres, Organisation maritime internationale.

PAASCH, H. (1885) : *De la quille à la pomme de mâ. Dictionnaire de marine en anglais, français et allemand illustré de nombreux dessins explicatifs*, 1^{re} édit., Anvers.

PAASCH, H. (1894) : *De la quille à la pomme de mâ. Dictionnaire de marine en anglais, français et allemand illustré de nombreux dessins explicatifs*. 2^e édit., Anvers, Paasch, Hamburg, Eckardt & Messtorff, Londres, Fisher.

PAASCH, H. (1901) : *De la quille à la pomme de mâ. Dictionnaire de marine en anglais, français et allemand illustré de nombreux dessins explicatifs*, 3^e édit., Anvers, Paasch, Hamburg, Eckardt & Messtorff.

PAASCH, H. CALLAMEL, P., MATTHIESEN, ?, BUDDE, A., MONTOJO, P. et G. ROMAIRONE (1908) : *From keel to truck, De la quille à la pomme du mâ, Vom Kiel zum Flaggenknoff, De quilla a perilla, Dalla chiglia al pomo dell'albero. Dictionnaire de marine anglais, français, allemand, espagnol, italien*. 4^e édition révisée et complétée, Paris, Augustin Challamel, Londres, George Philip & Son, David Nutt, Liverpool, Philip, Son & Nephew, Hamburg, Eckardt & Messtorff.

PAASCH, H., BATAILLE, L. et M. BRUNET (1937) : *From keel to truck, De la quille à la pomme du mâ, Vom Kiel zum Flaggenknoff, De quilla a perilla, Dalla chiglia al pomo dell'albero. Dictionnaire de marine anglais, français, allemand, espagnol, italien établi d'après l'ancien dictionnaire du capitaine Paasch [...] avec la collaboration du Bureau Veritas*, 5^e édit., 2 vol., Paris, Société d'éditions géographiques maritimes et coloniales, Londres, George Philip.

- RATCLIFF, R.E. (1983) : *Dictionnaire anglais-français des termes de marine = English-French dictionary of naval terminology. Revised and edited by Michaël A. Peck and Jean Maillot*, Paris, Technique et Documentation Lavoisier.
- RENIER, G. et J. VERSTRAETEN (1925) : *Arrimage des marchandises, jaugeage et classification des navires*, 2^e édit., Anvers, Association maritime belge (Bibliothèque de l'officier de la marine marchande belge).
- SIZAIRE, P. (1978) : *Les termes de marine*, 2^e édit., Paris, PUF, coll. «Que sais-je ?», n° 1479.
- SOÉ, G., DUPONT, J. et O. ROUSSIN (1910) : *Vocabulaire des termes de marine*, Paris, Le Yacht.
- Termdok2* (1992) : *Termdok2 on CD-Rom*, base de données sur cédérom, Stockholm, T.N.C., Swedish Centre for Technical Terminology et Walters Lexicon.
- Termium* (1996) : *Termium*, base de données sur cédérom, Ottawa, Gouvernement du Canada, Service de la traduction.
- VAN CAMPENHOUDT, M. (1994) : *Un apport du monde maritime à la terminologie notionnelle multilingue. Etude du dictionnaire du capitaine Heinrich Paasch De la quille à la pomme de mât (1885-1901)*, Paris, Université de Paris XIII, Thèse de doctorat, 2 vol.
- VANDENBERGHE, J.-P. et L.Y. CHABALLE (1978) : *Elsevier's nautical dictionary in six languages, English/American, French, Spanish, Italian, Dutch and German. Second, completely revised edition compiled and arranged on an English alphabetical basis*, Amsterdam, Oxford et New York, Elsevier Scientific Publishing Company.
- VANDENBERGHE, J.-P. et M. JOHNEN, M. (1994) : *Elsevier's nautical dictionary in English/American, French, Spanish, Italian, Dutch and German. Third completely revised edition*, Amsterdam, London, New York, Tokyo, Elsevier.
- VERHAEGE, L. (1994) : *Diccionario de la navegación marítima. Español - Inglés - Francés - Neerlandés basado en el vocabulario normalizado de la navegación marítima de la OMI*, Leuven, Wouters.
- VERSTRAETEN, J. (1913) : *Navire-école «Comte de Smet de Naeyer». Leçons élémentaires de matelotage, gréement et manœuvres*, Louvain, A. Uytenspruyt (manuscrit polycopié).
- WARTBURG, W. von (1950-1957) : *Französisches etymologisches wörterbuch*, 2^e édition, divers lieux et éditeurs, 25 vol. + supplément.

LES DÉNOMINATIONS EN LANGUE DE SPÉCIALITÉ : MONORÉFÉRENTIALITÉ ET PSEUDO-SYNONYMIE

Christine DURIEUX

Université de Caen, Caen, France

La ressemblance que produit l'idée générale fait donc les mots synonymes: et la différence qui vient de l'idée particulière qui accompagne la générale, fait qu'ils ne le sont pas parfaitement.
(Abbé Girard, 1718, *La Justesse de la langue françoise ou Différentes significations des mots qui passent pour être synonymes*)

INTRODUCTION

Ce thème de recherche, qui s'inscrit dans une relation entre la description de la langue, en particulier du lexique, et son application à la traduction a été motivé par l'observation d'erreurs de traduction imputables à la pseudo-synonymie, et cela tant en milieu pédagogique que dans le cadre de la traduction professionnelle.

Or, précisément dans les langues de spécialité, l'idée même de synonymie — future pseudo — semble d'emblée être exclue. En effet, les langues de spécialité font largement appel à des nomenclatures dont les éléments ont pour caractéristique majeure d'être monoréférentiels. Toutefois, bien que la notion même de langue de spécialité semble concerner exclusivement la traduction technique, à cet égard, il serait bien imprudent d'exclure la traduction littéraire. En effet, le texte littéraire n'est pas forme pure. En littérature romanesque, par exemple, le texte raconte une histoire, il met en scène des personnages qui évoluent dans un décor qui, même s'il est fictif, évoque des réalités relevant de diverses disciplines, empruntant ainsi constamment à différentes langues de spécialité.

À ce propos, introduire une typologie des textes, et notamment une distinction entre texte technique et texte littéraire, est tout à fait inappropriée, le critère de dichotomie étant très incertain, approximatif et flou. C'est pourquoi, le présent développement se situe dans le cadre de l'application à la traduction, lui conférant une validité pour tous les types de texte.

Dans l'énoncé ci-dessus, la qualification de la césure entre texte technique et texte littéraire fait appel aux adjectifs incertain, approximatif et flou, auxquels d'ailleurs il serait possible d'ajouter brouillé, indistinct, nébuleux, vaporeux, confus, indéfini, indiscernable, précaire, etc.; en effet, ces synonymes proposés par les dictionnaires pourraient s'appliquer «presque» indifféremment.

DÉFINITIONS

Langue de spécialité

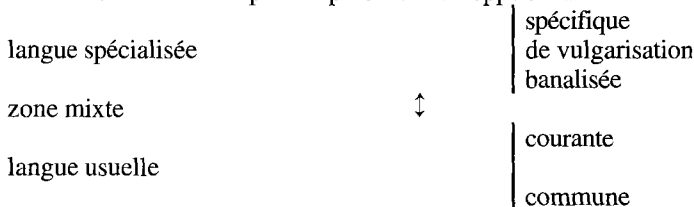
«Sous-système linguistique qui utilise une terminologie et d'autres moyens linguistiques et qui vise la non-ambiguïté de la communication dans un domaine particulier» (AFNOR 1990).

«On appelle langue de spécialité un sous-système linguistique tel qu'il rassemble les spécificités linguistiques d'un domaine particulier. En fait, la terminologie, à l'origine de ce concept, se satisfait très généralement de relever les notions et les termes considérés comme propres à ce domaine. Sous cet angle, il y a donc abus à parler de langue de spécialité, et vocabulaire spécialisé convient mieux» (Dubois 1994 : 440).

C'est cette dernière position qu'adopte P. Lerat (1995) qui considère qu'il ne peut exister à proprement parler des langues de spécialité parce que les activités humaines ne sont pas strictement cloisonnées ni cloisonnables. Rejetant de même la notion de technolecte, il suggère qu'il y aurait avantage à parler de *langue spécialisée*. Néanmoins, il précise : «Une langue spécialisée ne se réduit pas à une terminologie : elle utilise des dénominations spécialisées (les termes), y compris des symboles non linguistiques, dans des énoncés mobilisant les ressources ordinaires d'une langue donnée. On peut donc la définir comme l'usage d'une langue naturelle pour rendre compte techniquement de connaissances spécialisées» (1995 : 21).

Quoi qu'il en soit, spécialisée ou de spécialité, il s'agit d'une langue servant à véhiculer des connaissances spécialisées. À ce titre, elle peut être opposée à la langue usuelle. Toutefois, il semble que, dans une taxinomie des discours, il n'y ait pas opposition mais plutôt coexistence dans la continuité.

Aux fins de la présente étude, il importe de positionner la notion de langue de spécialité par rapport à la langue usuelle. Il apparaît pertinent ici de distinguer, à l'intérieur de la langue usuelle, entre *langue commune* normalement partagée par tous les membres de la communauté linguistique et *langue courante* enrichie d'emprunts à des champs d'expérience spécialisés et utilisée surtout par les classes socio-professionnelles favorisées (Galisson 1979) ou les plus cultivées. Une présentation schématique en strates superposées semble bien convenir pour le présent développement.



Cette représentation graphique est nécessairement heurtée pour montrer les principaux niveaux de langue. De fait, ni la langue usuelle, ni les langues de spécialité ne sont des ensembles homogènes bien délimités. Dans chacun d'eux, il existe plusieurs niveaux et registres. En réalité, il s'agit d'un continuum avec l'existence d'une zone mixte intermédiaire, passage obligé de l'injection de termes et de phraséologismes spécialisés dans

la langue courante et de l'aspiration de termes appartenant à la langue usuelle dans des langues de spécialité, de même que du retour dans la langue courante de termes initialement empruntés à la langue usuelle par des domaines spécialisés, après transformations résultant de leur emploi dans une ou plusieurs langues de spécialité.

On observe que les mouvements entre langue usuelle et langue spécialisée se font à double sens. Or, le transit des unités lexicales ne se fait pas par sauts quantiques et il n'existe pas de bande interdite entre les niveaux de langue comme celle qui sépare la bande de valence de la bande de conduction d'un atome. Au contraire, les migrations terminologiques se font en continu, passant de la langue usuelle à la langue spécialisée et inversement par une zone mixte où des unités de la langue usuelle se chargent de valeurs spécialisées et où des unités d'une langue spécialisée, étant devenues tellement banalisées, sont prêtes à s'intégrer à la langue courante.

De plus, cette zone mixte intermédiaire subit une évolution dans le temps. Elle peut d'abord se démarquer de la langue spécialisée, d'une part, et de la langue usuelle, d'autre part, acquérant un statut indépendant parfois éphémère mais qui, en tout état de cause, n'a pas pour mission de perdurer. Ensuite, elle assure la transition entre langue spécialisée et langue usuelle avant d'être l'agent support de l'intersection entre les deux. On observe finalement un recouvrement partiel de la langue usuelle courante par la langue spécialisée banalisée au point qu'il soit impossible de distinguer l'une de l'autre, sinon dans la nuance de tel ou tel usage.

Ce phénomène de fusion diachronique garantit, en quelque sorte, l'existence d'un continuum dans l'échelle des niveaux de langue.

Pseudo-synonyme

Pseudo- : racine provenant du mot grec *pseudos*, signifiant mensonge délibéré ou par erreur (trompeur), d'où l'idée de faux ou de faussement appelé, s'appliquant au lexème ainsi préfixé.

Ce préfixe a été choisi pour la notion de «trompeur» qu'il véhicule. De fait, avec les pseudo-synonymes, on est en présence d'une illusion de synonymie, de nature à induire le traducteur en erreur.

Cette notion est à distinguer de celles de parasyndrome et de quasi-synonyme qui sont des synonymes incomplets. Se fondant sur l'absence de synonyme parfait ou total, les termes en relation de parasyndromie ou de quasi-synonymie sont censés partager un grand nombre de traits pertinents en commun, la différenciation se faisant au niveau de la distribution, c'est-à-dire des conditions discursives dans lesquelles ils sont utilisés.

Dans la terminologie de la traductologie, un pseudo-synonyme est un faux synonyme ou plutôt une unité lexicale faussement considérée comme synonyme, en ce sens qu'il ne présente pas toutes les caractéristiques d'un synonyme. En particulier, dans le présent développement, il s'en distingue par le fait qu'il est substituable à une autre unité lexicale dans certains énoncés, mais pas dans d'autres, avec la particularité que la césure se fait à un point de passage à un niveau de spécialisation plus élevé de la langue. Ce point de passage peut se situer entre le niveau de la langue courante et celui de la langue spécialisée

banalisée, ou à l'intérieur d'une langue de spécialité entre un niveau d'emploi banalisé et un niveau plus finement spécialisé.

Dans ses travaux, Y. Gentilhomme (1994 : 387) appelle «termes larvés [...] des mots qui, à première vue, semblent appartenir au vocabulaire commun, mais qui, dans la discipline considérée, acquièrent un contenu particulier, non explicitement défini; on ne le saisit qu'à la suite d'une pratique plus ou moins prolongée de la discipline».

À titre d'illustration, l'introduction du présent article rappelle que le critère de distinction entre texte technique et texte littéraire est notamment qualifiable d'incertain, approximatif et flou. Mais il ne faudrait pas s'y tromper, on est ici en présence de termes larvés évoqués ci-dessus, ou de pseudo-synonymes pour le traducteur. En effet, en intelligence artificielle, par exemple, il y a lieu d'établir une distinction rigoureuse entre ces trois adjectifs.

«Une connaissance *incertaine* est une connaissance pour laquelle l'information concernant l'un des attributs est inconnue (exemple : la cause de la disparition des dinosaures). Une connaissance *approximative* est une connaissance pour laquelle l'information concernant l'un des attributs est connue mais seulement de manière approximative (exemple : la datation de la disparition des dinosaures). Un prédicat *flou* est un prédicat qui peut fournir des valeurs de vérité autres que "vrai" ou "faux" (c'est-à-dire allant du vrai au faux en passant par le quasiment vrai, le presque vrai, l'assez vrai, le plus vrai que faux, le ni vrai ni faux, le plus faux que vrai, l'assez faux, le presque faux et le quasiment faux» (Ozman 1996 : 116).

PSEUDO-SYNONYMES ET TRADUCTION

Les pseudo-synonymes ainsi définis posent un problème particulier en traduction. En effet, le traducteur, qui peut être familiarisé avec une discipline technique mais dont on ne peut s'attendre qu'il soit expert de cette discipline, risque de ne pas être en alerte et de voir une synonymie, habituelle dans la langue usuelle, là où il y a différenciation, voire opposition, en langue de spécialité.

Passage de langue usuelle à langue de spécialité

Plus le domaine de spécialité est proche des préoccupations quotidiennes du grand public, plus le sous-domaine touche à la vie courante, et plus le danger est grand pour le traducteur de ne pas reconnaître dans des unités lexicales appartenant à la langue usuelle l'emploi de véritables termes relevant d'une langue de spécialité. C'est le cas, par exemple, du domaine de l'écologie, sous-domaine de la gestion des déchets, dans lequel les informations sont abondantes, récurrentes, d'accès facile parce que concrètes et ayant trait à des aspects de la vie de tous les jours. En outre, la langue de spécialité emprunte de nombreux éléments du lexique de la langue usuelle, ce qui accroît la difficulté de discerner entre (1) unités lexicales courantes avec les acceptions approximatives et polysémiques habituelles et (2) termes spécifiques avec la couverture conceptuelle strictement bornée correspondante tendant vers la monoréférentialité.

En tant que membre d'une communauté cultivée et informée, le traducteur croit pouvoir gagner du temps en faisant l'économie d'une recherche documentaire rigoureuse. Il

se sent en confiance et se contente de mobiliser les connaissances thématiques dont il dispose. Il risque alors de considérer comme étant interchangeables des termes tels que *réemploi* et *réutilisation*. De fait, en langue courante, on est tenté de voir là un cas type de synonymie, tentation d'ailleurs renforcée par une éventuelle étude étymologique. En effet, dans les deux cas, la morphologie est de même type : préfixe *ré* indiquant une répétition devant un déverbal usuel *emploi* et *utilisation*.

En outre, le dictionnaire Larousse donne, pour ces deux substantifs, les définitions suivantes :

EMPLOI : Action ou manière d'employer; utilisation d'une chose.

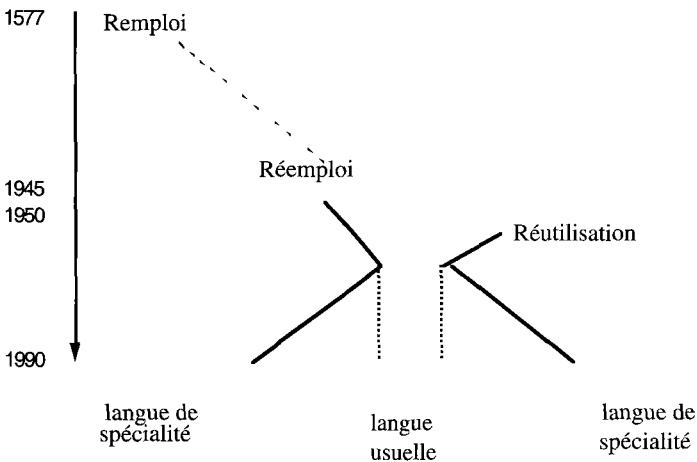
UTILISATION : Action d'utiliser.

De prime abord, on est tenté de relever une grande similitude entre les deux substantifs, l'un figurant même (comme synonyme ?) dans la définition de l'autre.

En revanche, si *réutilisation* n'est pas répertorié en entrée des dictionnaires de langue tels que Robert et Larousse, on y trouve *réemploi* et sa variante *emploi* avec pour définition : Achat d'un bien avec le produit de la vente ou de l'indemnisation de la perte d'un autre bien. Ce terme est attesté dès 1577 dans sa forme *emploi* pour désigner l'acquisition de biens mobiliers ou immobiliers avec les fonds provenant de la cession d'un bien dotal.

Ainsi donc, ces deux substantifs apparemment assimilables l'un à l'autre dans la langue usuelle actuelle ont déjà fait l'objet d'usages divergents. Ils procèdent de voies diachroniques radicalement différentes. *Remploi*, avec son appartenance à une langue de spécialité — celle du droit dotal — a une grande antériorité par rapport à *réemploi* qui n'apparaît qu'en 1945. Quant à *réutilisation*, ce mot est de formation encore plus récente.

Dans la langue usuelle, *réemploi* peut concerner aussi bien un objet inanimé qu'une personne, alors que *réutilisation* porte seulement sur des objets inanimés, sauf usage métaphorique.



Aujourd'hui, dans la langue spécialisée de la valorisation des déchets, ces deux termes ont des significations très nettement distinctes, voire opposées.

- le *réemploi* prolonge la durée de vie d'un produit par un second emploi analogue au premier. Par exemple, consignées, les bouteilles peuvent être à nouveau remplies après nettoyage.

- la *réutilisation* consiste à affecter un déchet à un usage différent de son usage initial ou à l'inclure dans la fabrication d'un nouveau produit. Par exemple, les pneus de voiture usagés peuvent servir à protéger la coque des bateaux.

Cette illusion de synonymie, liée à la carence des connaissances thématiques du traducteur, est fréquente et peut se manifester chaque fois qu'une langue de spécialité aspire des unités lexicales de la langue courante et, en les absorbant, en précise ou en modifie la surface conceptuelle correspondante. Alors que ce processus tend à conférer un caractère monoréférentiel aux termes ainsi intégrés dans les langues de spécialité, qui réfèrent à des réalités (ou *realia*) particulières et exclusives, le traducteur risque de ne pas en prendre conscience et de s'en tenir à ce qu'il sait des contenus polyvalents et des emplois variables de ces unités lexicales en langue usuelle. Les conséquences dans la production de traductions peuvent être extrêmement lourdes, puisque cette illusion peut donner lieu à de graves contre-sens.

En langue de spécialité

En langue de spécialité, dans un domaine strictement borné, sans aucune référence à la langue usuelle, une pseudo-synonymie peut se manifester et poser des problèmes de traduction en fonction du niveau de spécialisation du texte. Plus ce niveau est élevé, plus la différenciation des termes est fine et significative. En outre, plus le domaine de spécialité progresse et acquiert une maturité, plus sa terminologie se précise et se différencie, tendant ainsi vers la monoréférentialité.

«La terminologie de la théorie de E. Wüster affirme la monoréférentialité du terme scientifique ou technique; c'est-à-dire que, dans un domaine étroitement défini, le terme ne désigne qu'une classe d'objets, de qualités ou de procès : le référent est unique, par l'intermédiaire de la notion, insérée dans un système notionnel» (Dubois, 1994 : 309).

Par exemple, dans le domaine de la communication d'entreprise, qui résulte d'une prise de conscience relativement récente des entreprises et qui, de ce fait, donne lieu à des activités nouvelles et en constante évolution, la terminologie se développe parallèlement, générant une pseudo-synonymie.

À côté de la réclame du milieu de ce siècle, qui s'est commuée en publicité, avec ses règles et ses codes, sont apparues de nouvelles formes de communication par lesquelles l'entreprise fait valoir son image et renforce sa notoriété : le mécénat, le parrainage, le patronage et le sponsoring.

Depuis 1973, la BNP fait figurer son logo sur les bâches de fond de court tout au long du tournoi de tennis de Roland-Garros. Ce faisant, la banque fait-elle oeuvre de mécène, parraine-t-elle le tournoi, le patronne-t-elle ou le sponsorise-t-elle ? Ou encore,

s'inscrit-elle dans un partenariat ? Ces différents termes sont-ils interchangeables, c'est-à-dire sont-ils synonymes, ou bien sont-ils seulement quasi-synonymes, para-synonymes, ou sont-ils pseudo-synonymes ?

Seule une étude rigoureuse de l'emploi de ces termes par les spécialistes de la discipline peut permettre d'apporter une réponse à cette question. De fait, la consultation de dictionnaires n'apporte aucun éclaircissement.

Grand dictionnaire Larousse Encyclopédique :

MÉCÉNAT - Protection accordée aux lettres, aux arts et aux sciences en qualité de mécène.

MÉCÈNE - (du nom du favori de l'empereur Auguste, Mécène). Celui qui protège des lettrés, des savants, des artistes, en les aidant par de puissants moyens financiers.

PARRAINAGE - Action de parrainer, qualité, relations, fonctions de parrain.

PARRAINER - Servir de parrain à.

PATRONAGE - Protection exercée par un patron, par un personnage puissant.

Il est inutile de chercher à multiplier les définitions puisées dans des dictionnaires de langue; toutefois, la consultation du *Dictionnaire historique de la langue française* peut permettre d'aider à mieux cerner la valeur actuelle de ces termes en se référant à leur origine.

MÉCÈNE - est tiré (1626) du latin Maecenas, nom d'un chevalier romain du I^{er} siècle avant J.C., descendant d'une noble famille étrusque, conseiller d'Auguste et protecteur des belles lettres (en particulier de Virgile et d'Horace). Naturellement, dans l'Antiquité comme en France, il existait des mécènes et un mécénat avant la lettre, pratiqué par les souverains, les prêtres puis, à partir des XIII-XIV^{ème} siècle par des laïcs [...] Depuis la première guerre mondiale, il convient de souligner le rôle des grands collectionneurs, notamment américains, et le rôle de l'État et de l'industrie.

MÉCÉNAT - (1867) D'abord terme d'Antiquité, a suivi l'évolution de mécène. L'usage de la commandite publicitaire, le succès de l'américanisme *sponsor* et de ses dérivés donne à *mécène* et à *mécénat* la possibilité de nouveaux emplois.

PARRAINAGE - (1200) D'abord écrit parrinnaige, a remplacé compérage. Le mot a été repris sous sa forme moderne au XIX^{ème} siècle pour désigner la qualité, la fonction de parrain [...] Parrainage correspond à l'appui moral prêté par une personne d'autorité à une oeuvre (1935) [...] *Parrainage* et *parrainer*, en relation avec *mécénat* et *mécène*, ont été proposés récemment pour remplacer l'anglicisme *sponsor* et ses dérivés.

PATRONAGE - Depuis le début du XIV^{ème} siècle, *patronage* est employé au sens général de «protection» et, par une métonymie ultérieure, il désigne l'organisation qui apporte une aide à des personnes démunies (1859).

PATRONNER - (1501) «Couvrir de son crédit, de sa protection», n'est plus attesté avant 1611, puis semble disparaître. Il est repris au XIX^{ème} siècle (1839).

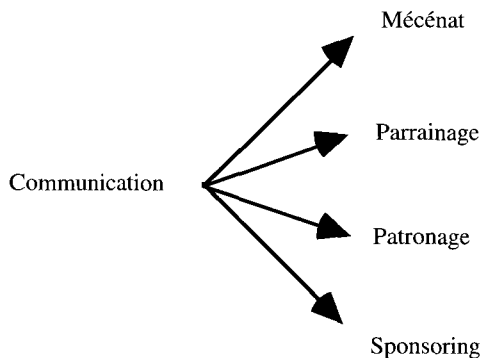
SPONSOR - Est un emprunt (1954) à un mot anglais (XVII^{ème} siècle) signifiant «parrain», «répondant, caution», qui a pris aux États-Unis le sens de «bailleur de fonds» (1931),

«commanditaire d'émissions de radio, de télévision». Le mot est emprunté au latin classique *sponsor* «répondant, caution» et en latin ecclésiastique «parrain d'un néophyte». Employé d'abord dans le domaine des sports, cet anglicisme désigne une personne ou un organisme qui soutient financièrement une entreprise ou un club sportif, équivalent du français commanditaire ou, avec d'autres connotations, de mécène.

SPONSORING - (1972) A surtout cours dans le monde du sport. Il est concurrencé en français par *sponsorisation* (1980). Critiquée, cette série d'anglicismes est d'usage fréquent, mais *parrainage* semble concurrencer *sponsorisation*.

L'étude diachronique de la langue tend à mettre en évidence une convergence entre ces quatre termes en signalant des possibilités de substitution entre sponsor et mécène, entre parrainage et parrainer, mécénat et mécène, et sponsor, entre parrainage et sponsorisation. En outre, dans les définitions de ces quatre termes, on retrouve un lien isotopique assuré, notamment, par les termes protecteur, protection, parrain. Il semble donc qu'il y ait bien un chevauchement entre les surfaces conceptuelles de ces quatre termes, chevauchement qui d'ailleurs justifie leur appartenance à un même sous-sous-domaine appelé la publicité par l'événement (Decaudin 1995) encore dénommée communication institutionnelle d'entreprise (Dambon 1993).

On pourrait déjà présenter un premier réseau sémantique avec la relation *sorte de*. Les quatre termes étudiés se trouvent en relation d'isonymie, c'est-à-dire à un même niveau d'hyponymie par rapport au générique, communication par l'événement. On peut dire également que ces quatre termes sont cohyponymes.



Au seul niveau linguistique, c'est-à-dire sans consulter les spécialistes du domaine, le terme *sponsoring* considéré comme un anglicisme ou comme un américanisme est d'emblée dénoncé et, en dehors de toute analyse sémique ou componentielle, il est a priori envisagé de le remplacer par des termes à consonance jugée plus exclusivement française tels que parrainage, par exemple. Lors de la première Université d'automne en terminologie (Rennes 2 - 1992), Loïc Depecker, de la Délégation générale à la langue française, déclarait : «Pour le terme *sponsor*, nous trouvons les termes *parrain* et *parraineur*; nous pensons qu'ils peuvent fonctionner et les imposer. Je trouve très bien le fait que, si ces termes ne plaisent pas au public, ou s'ils ne conviennent pas au message à faire passer, on voit arriver des périphrases, des phraséologismes tels que : *avec le soutien de, partenaire*

officiel de, etc.». L'inconvénient de ce type de raisonnement est qu'il porte uniquement sur les dénominations, sans référence aux réalités ainsi dénommées.

Après avoir vu les convergences, chevauchements et recouvrements proposés par les lexicographes, qui laissent une impression générale d'amalgame et de substituabilité, il n'est pas inutile de voir ce qu'en disent les spécialistes du domaine à cet égard. On constate qu'ils utilisent ces quatre termes pour désigner des réalités qui, certes, gravitent dans le même champ d'activité — la communication d'entreprise par l'événement — mais qui désignent des actions radicalement différentes dans leurs objectifs, leurs intentions, leurs cibles, leur champ d'application, la durée de l'action, la rationalité de la démarche et le mode de mise en valeur (Dambon 1993).

La confusion terminologique est telle, dans ce domaine, que P. Dambon (1993) juge opportun de consacrer la première partie de son ouvrage sur le mécénat et le sponsoring à préciser les définitions des termes mécénat, sponsoring, parrainage, patronage et partenariat et à expliquer ce qui les différencie. Nous retiendrons donc les définitions auxquelles il apporte sa caution.

MÉCÉNAT : N.m. (du lat. Maecenas, Ministre d'Auguste) Protection accordée aux lettres, aux sciences et aux arts.

SPONSORING : N.m. (de l'angl. to sponsor : cautionner). Aide financière apportée à un sport, à des fins publicitaires, par une firme ou un secteur commercial.

Le terme *mécénat* semble correspondre à un concept connu, défini et cerné de longue date. Toutefois, avec l'arrivée des nouvelles formes d'action de communication d'entreprise, le mécénat en vient à désigner «une forme d'expression, une rencontre, une proposition de dialogue : manière d'exprimer sa prise de responsabilité civique, communautaire, culturelle» (Vescia 1987 : 34). Ainsi, le mécénat d'entreprise oeuvre pour le développement d'activités artistiques, et permet en quelque sorte à l'entreprise de s'acquitter de ses devoirs à l'égard du grand public. Bien sûr, l'entreprise cherche à faire parler d'elle, mais il n'y a pas dans l'action de mécénat la recherche de retombées commerciales directes comme dans une action publicitaire ou une action de sponsoring. Par ailleurs, le mécénat intervient dans le domaine culturel ou humanitaire alors que le *sponsoring* intervient principalement dans le domaine sportif. «Le sponsoring est un outil de communication permettant de lier directement une marque ou une société avec un événement (sportif) attractif pour un public donné» (Sahnoun 1986 : 18). Décaudin (1995) propose un autre type de différenciation entre ces deux termes.

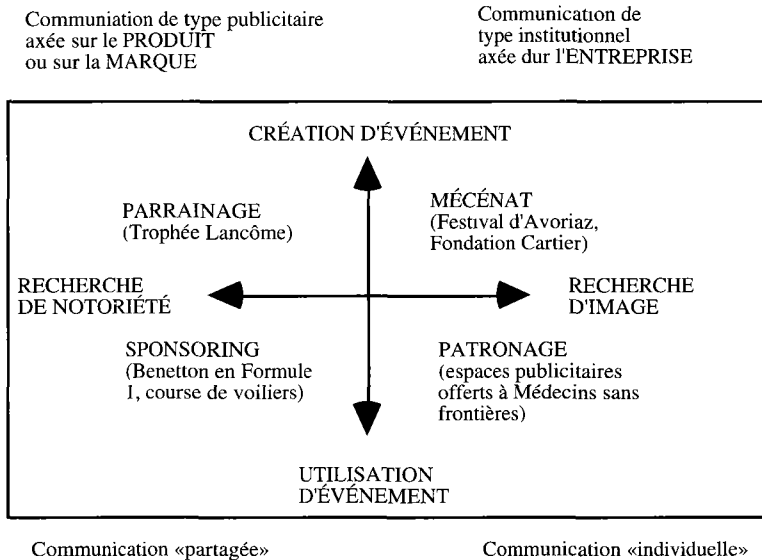
Le *sponsoring* qualifie une action de publicité par l'événement à connotation commerciale dont l'effet est attendu à court terme; ceci implique une présence voyante sur l'événement et dans son exploitation médiatique; le *mécénat* qualifie une action de publicité par l'événement orientée vers une amélioration d'image dont l'effet est attendu à moyen (ou long) terme; ceci implique une présence discrète sur l'événement et dans son exploitation médiatique.

Les critères de différenciation de ces deux notions portent ici principalement sur le délai de manifestation attendue des retombées et l'exposition publicitaire de l'entreprise.

Face à ces deux notions, manifestement bien différenciées, le *parrainage* apparaît comme une forme de sponsoring appliqué à un événement le plus souvent créé par l'entreprise à cet effet.

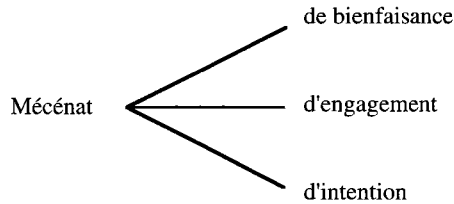
«Le terme de patronage, "protection accordée par un homme puissant à un homme d'état inférieur" traduit l'inégalité de puissance entre les contractants. La notion de parrainage ne véhicule pas ce déséquilibre et se limite à traduire la promotion d'un événement permise par l'intervention de l'annonceur.» (Grégory 1984 : 168). Le *patronage* serait donc une action visant à apporter un appui à une personne, à une institution ou à un événement, sous forme d'encouragement sans qu'il y ait nécessairement soutien financier. D'une manière générale, le patronage intervient pour soutenir des manifestations culturelles ou liées à des grandes causes. Il n'y a pas de création d'événement, mais plutôt un appui discret apporté à un événement déjà existant.

Simon Loutrel (1985 : 43) établit une distinction entre ces quatre termes selon que l'entreprise crée l'événement ou au contraire utilise l'événement par l'intermédiaire duquel elle souhaite communiquer. D'après ce qui précède, il est possible de positionner ces quatre termes, non plus en fonction de leur étymologie et de leur dynamique diachronique, mais en fonction du contenu sémantique que leur confèrent les spécialistes du domaine. Les critères de différenciation sont alors multiples.

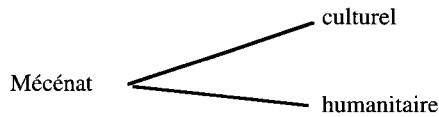


La consultation d'ouvrages (et il sont nombreux !) publiés par les spécialistes du sujet a donc permis déjà de positionner ces quatre termes les uns par rapport aux autres. En outre, cette recherche documentaire a permis de mettre en évidence un autre niveau de différenciation. En effet, il existe diverses formes de mécénat comme il existe diverses formes de sponsoring, l'apparition de ces sous-catégories suivant des critères différents. Il est ainsi possible de tracer de nouveaux réseaux sémantiques.

En fonction de ses objectifs



En fonction de son champ d'application

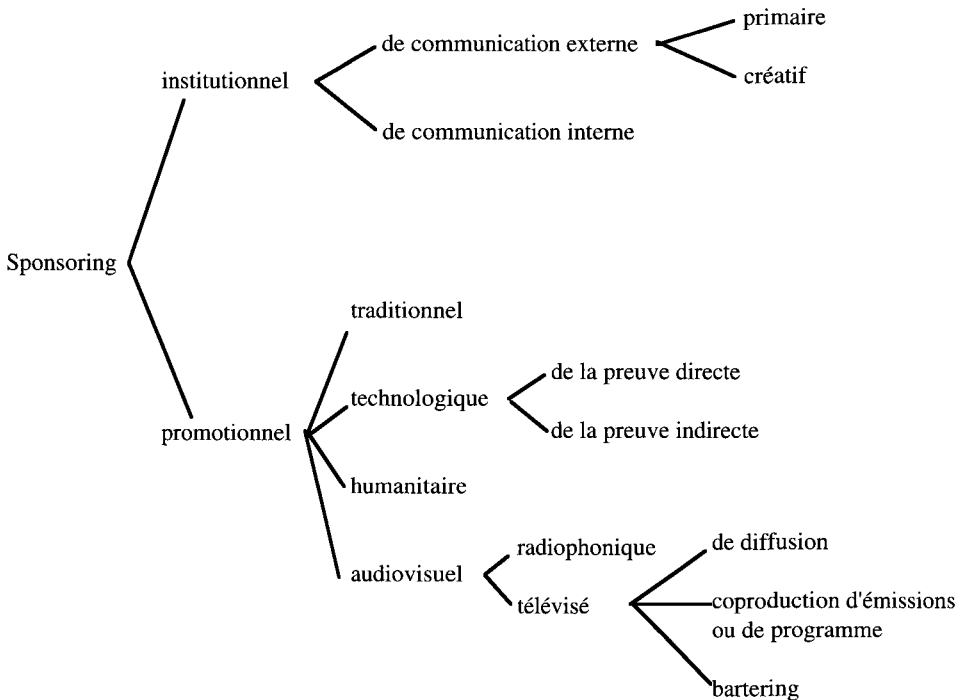


Toutefois, les cooccurrents de mécénat ne se limitent pas à ces quelques termes. La recherche documentaire permet de relever d'autres collocations. Par exemple :

Le *mécénat associé* consiste à faire participer les salariés à une action le plus souvent humanitaire pouvant être identifiée comme une action de l'entreprise, en les associant aux choix de l'opération et en les faisant coopérer. Il s'agit, en fait, d'un *mécénat de proximité* (Dambron 1993). Pour comprendre ce dont il s'agit, il est bien inutile d'analyser les composants lexicaux de ces termes, c'est-à-dire de s'intéresser à l'amont; il est beaucoup plus efficace de s'intéresser à l'aval avec un exemple d'application sur le terrain. À cet égard, on peut citer notamment le cas de la société Apple qui a équipé d'ordinateurs Macintosh l'hôpital des enfants malades de Garches, afin de permettre aux enfants de travailler, de se distraire et de réaliser un journal interne. Des cadres informaticiens de la société viennent les former gracieusement. Pour que l'opération fonctionne, le constructeur fournit les ordinateurs et les programmes adaptés, et indique les coordonnées d'un technicien bénévole qui pourra intervenir en cas d'éventuels problèmes. Il est clair que c'est la connaissance de la réalité qui permet de comprendre et d'utiliser le terme, et non l'analyse de la motivation de celui-ci.

Il est possible de procéder de même pour le sponsoring. D'après les spécialistes de la communication d'entreprise, il y a effectivement lieu de distinguer entre le *sponsoring institutionnel*, qui s'inscrit dans le cadre d'une communication institutionnelle, et le *sponsoring promotionnel* qui relève davantage d'actions de marketing. En outre, à l'intérieur du sponsoring institutionnel, on relève une forme de communication externe censée avoir des répercussions sur la perception à l'extérieur de l'activité de l'entreprise, de ses produits et de ses services, et la communication interne ciblée sur son propre personnel. À l'intérieur des opérations de sponsoring promotionnel, il y a lieu de distinguer entre *sponsoring traditionnel*, c'est-à-dire qui est passé dans les habitudes et dans l'usage, *sponsoring technologique*, *sponsoring humanitaire* et *sponsoring audiovisuel*. Le *sponsoring technologique* — par exemple, Elf en Formule 1 — est presque une forme de transfert de technologie. L'entreprise développe un produit de pointe et le fait intervenir dans son action de sponsoring : on parle alors de sponsoring de la preuve directe.

L'entreprise sponsor peut aussi faire bénéficier le sponsorisé des techniques qu'elle a mises au point et qu'elle continue à affiner, bénéficiant elle-même ainsi par son action d'un effet de laboratoire : c'est ce qu'on appelle le sponsoring de la preuve indirecte. Le *sponsoring humanitaire* est une forme d'action publicitaire associée à une opération caritative. C'est le cas par exemple d'Evian qui faisait figurer sur ses bouteilles d'eau minérale son engagement à verser, pour chaque bouteille vendue, une certaine somme à la Croix Rouge. Le *sponsoring audiovisuel* se définit le plus généralement par la présence notoire à la télévision d'une entreprise, de sa marque ou de ses produits en dehors des écrans publicitaires normalement programmés. C'est un mode d'action surtout utilisé par les entreprises qui n'ont pas le droit de faire des spots télévisés classiques : chaînes de distribution, par exemple. Bien entendu, le sponsoring audiovisuel regroupe le sponsoring télévisé et le sponsoring radiophonique. À l'heure actuelle, le *sponsoring télévisé* est le plus répandu; il est aussi multiforme : *sponsoring de diffusion* avec présence du nom de l'entreprise dans le programme, dans le générique ou dans les bandes annonces, *coproduction d'émissions* et coproduction des programmes, l'entreprise étant partenaire dans le financement d'émissions et de programmes et, à ce titre, figurant non seulement au générique mais éventuellement à l'intérieur même de l'émission, et *bartering*, c'est-à-dire troc, consistant pour l'entreprise à proposer à une chaîne de télévision un programme tout fait en échange d'écrans publicitaires. C'est ce dernier type d'opérations qui a motivé la formation du terme anglais *soap-opera*, apparu aux États-Unis dans les années 60 quand les fabricants de lessive Procter & Gamble et Unilever ont proposé des feuillets à des chaînes de télévision en échange d'écrans publicitaires. Ce rapide balayage documentaire peut donner lieu au réseau sémantique suivant :



Dans ce réseau, à tous les niveaux, la relation est *forme de*.

Ce réseau ne saurait être considéré comme exhaustif quant aux cooccurrents du terme *sponsoring*. En effet, selon le degré d'implication de l'entreprise dans l'opération de *sponsoring*, il y a lieu de distinguer entre *sponsoring primaire* qui consiste pour l'entreprise à acheter de l'espace publicitaire en faisant figurer son nom sur un voilier, une automobile, le maillot ou le dossard d'un joueur, etc. et le *sponsoring créatif* qui consiste pour l'entreprise à apporter sa contribution au développement d'une activité sportive, culturelle, ou autre.

CONCLUSION

Il apparaît clairement qu'une approche linguistique de la terminologie ne permet pas de rendre compte du contenu « technique » des termes en présence. On remarque que c'est moins la définition qui permet de comprendre la réalité désignée que l'exemple de manifestation réalisée qui permet d'inférer les caractéristiques et traits distinctifs des différents concepts. À l'issue de ce rapide tour d'horizon, il paraît bien dérisoire de s'acharner contre le terme *sponsoring*. Le *Journal Officiel de la République Française* du 3 avril 1982 précise que ce terme est à proscrire et à remplacer par *parrainage* avec la définition suivante :

PARRAINAGE : (N.m.) Soutien financier apporté par un commanditaire en contrepartie du surcroît de notoriété qu'il en escompte. Anglais : *sponsoring*.

Dans son communiqué du 13 juin 1985, l'Académie française stipule à propos de *sponsoring* : « Cet anglicisme est à éviter et doit être remplacé par les verbes : *patronner*, *financer*, *parrainer*, *commanditer*. On dira *commanditaire* à la place de *sponsor*, et *parrainage* à la place de *sponsorisation* ou de *sponsoring* ».

Or, il est rare que les termes imposés par des autorités extérieures aux disciplines dans lesquelles ils doivent être utilisés soient effectivement adoptés. Perçus comme des unités exogènes, ils ne sont généralement pas repris par les spécialistes. En revanche, ces derniers n'hésitent pas à faire preuve d'imagination, comme en témoigne la prolifération d'appellations largement diffusées à propos de la Coupe du Monde de Football qui aura lieu en France en 1998 : La Poste, *opérateur* officiel; Manpower, *réseau* officiel; Eurocard-Mastercard, *carte* officielle; SFR, *partenaire* officiel; Danone, *fournisseur* officiel du Mondial 98.

Cette lutte désespérée, et résolument contre-productive, ne contribue manifestement qu'à accroître la confusion et à nuire à la communication. En effet, en recommandant la substitution d'un terme par un autre, elle crée une convergence lexicale là où les spécialistes du domaine voient une différenciation. Elle renforce la pseudo-synonymie, source de tant d'erreurs en traduction, là où la monoréférentialité, existante dans l'usage parmi les spécialistes, devrait être maintenue voire soutenue. En outre, en allant à l'encontre de l'usage, elle ne peut que nuire à la clarté et à l'efficacité de la communication. Elle vise à détruire le consensus, pourtant indispensable, sur lequel les spécialistes d'une discipline peuvent fonder une communication efficace et économique, tant entre eux que dans les relations entre spécialistes et non-spécialistes.

RÉFÉRENCES

1. À propos de terminologie

AFNOR (1990) : *Terminologie*, Norme ISO 1087, Paris.

DUBOIS, Jean et coll. (1994) : *Dictionnaire de linguistique et des sciences du langage*, Paris, Larousse.

DUBUC, Robert (1992) : *Manuel pratique de terminologie*, Québec, Linguatex.

GALISSON, Robert (1979) : *Lexicologie et enseignement des langues*, Paris, Hachette.

GALISSON, Robert et Daniel COSTE (1976) : *Dictionnaire de didactique des langues*, Paris, Hachette.

GENTILHOMME, Yves (1994) : «Termes et symboles discours hétérogènes. Quelques hypothèses sémiologiques», Clas André et Pierrette Bouillon (dir), *TA-TAO : Recherches de pointe et applications immédiates*, Beyrouth, FMA et AUPELF-UREF.

GOUADEC, Daniel (1990) : *Terminologie - Constitution des données*, Paris, Afnor Gestion.

GOUADEC, Daniel (1993) : *Terminologie et Terminotique*, Paris, La Maison du Dictionnaire.

LERAT, Pierre (1995) : *Les langues spécialisées*, Paris, PUF.

OTMAN, Gabriel (1996) : *Les représentations sémantiques en terminologie*, Paris, Masson.

2. À propos de mécénat, sponsoring,...

BENVENISTE, François et Sylvère PIQUET (1988) : *Pratique du parrainage*, Paris, Vuibert entreprise.

BIOJOUT, Pascal (1983) : *Le sponsoring : analyse économique du comportement des entreprises en matière de parrainage sportif*, Faculté de droit et de sciences économiques de l'Université de Limoges.

BREBISSON, Guy de (1986) : *Le Mécénat*, Paris, PUF, coll. «Que sais-je ?».

Collectif (1982) : *À la découverte du sponsoring*, Paris, CECOD, Livrets du commerce moderne.

Collectif (1989) : *Le Mécénat dans l'histoire*, Actes du colloque du 21 mars 1989, Fondation Électricité de France, Paris, Éditions Sodel.

- DAMBRON, Patrick (1991) : *Sponsoring et politique de marketing*, Paris, Les Éditions d'Organisation.
- DAMBRON, Patrick (1993) : *Mécénat et sponsoring*, Paris, Les Éditions d'Organisation.
- DÉCAUDIN, Jean-Marc (1995) : *La Communication Marketing*, Paris, Economica.
- DOMINITZ, Jean-Charles et Solange TOCHON (1988) : *La Communication promotionnelle*, Paris, Les Éditions d'Organisation.
- GARRABOS, Christian (1987) : *Sports, sponsors et communication*, Paris, Sports Entreprises Médias Partenaires.
- GODELWSKI-SEGRESTAN, Stéphane (1991) : *Mécénat d'entreprise et stratégie*, Paris, Dunod.
- GRÉGORY, Pierre (1984) : «Sponsoring et Mécénat, instruments de communication institutionnelle», *La Revue Française de Gestion*, sep-oct.
- LOUTREL, Simon (1985) : «Le sponsoring, ça se gère», *Stratégies*, 8 juillet.
- PERRIN, Alain-Dominique (1986) : *Le Mécénat français*, Paris, Fondation Cartier.
- PIQUET, Sylvère et U.D.A (1985) : *Sponsoring et mécénat, la communication par l'événement*, Paris, Vuibert gestion.
- PLAT-PELLEGRINI, Véronique et Alain CORNEC (1987) : *Sponsoring, le parrainage publicitaire*, Paris, J. Delmas et C^{ie}.
- PROUST, Odile et ADMICAL (1982) : *On demande entreprises mécènes*, Paris, Chotard.
- REGOUBY, Christian (1988) : *La Communication globale*, Paris, Les Éditions d'Organisation.
- SAHNOUN, Pierre (1986) : *Le Sponsoring, mode d'emploi*, Paris, Chotard.
- SZYBOWICZ, André et Sylvie MAGISTRALI (1988) : *Sponsoring et mécénat, l'exemple de l'environnement*, Paris, Les Éditions d'Organisation.
- VESCIA, Rémo (1987) : *Le Mécénat : art de la communication, communication de l'art*, Paris, Economica.
- WEIL, Pascale (1990) : *Communication oblige !*, Paris, Les Éditions d'Organisation.

LA GRAMMAIRE CATÉGORIELLE COMBINATOIRE APPLICATIVE APPLIQUÉE AU FRANÇAIS

Ismail BISKRI^(1,2), Jean Pierre DESCLÉS⁽²⁾, Christophe JOUIS^(3,2)

(1) *Laboratoire de l'Analyse Cognitive de l'Information, Université du Québec à Montréal, Canada;* (2) *CAMS-LaLIC, Université de Paris Sorbonne, France;*

(3) *IDIST, Université de Lille 3, France*

1. LE MODÈLE DE LA GRAMMAIRE APPLICATIVE ET COGNITIVE

La Grammaire Applicative et Cognitive (Desclés 1990) est une extension de la Grammaire Applicative Universelle (Shaumyan 1987; Desclés 1990). Elle postule trois niveaux de description des langues :

- a- le niveau phénotypique (ou le phénotype) où sont représentées les caractéristiques particulières des langues naturelles (par exemple, l'ordre des mots, les cas morphologiques, etc.). Les expressions linguistiques de ce niveau sont des unités linguistiques concaténées, la concaténation est notée par : 'u₁-u₂-...-u_n';
- b- le niveau génotypique (ou le génotype) où sont exprimés les invariants grammaticaux et les structures sous-jacentes aux énoncés du niveau phénotypique. Le niveau génotypique est structuré comme un langage formel appelé «langage génotype»; il est décrit par une grammaire appelée «grammaire applicative»;
- c- le niveau cognitif où sont représentées les significations des prédicats lexicaux par des schèmes sémantico-cognitifs.

Les trois niveaux font appel à des formalismes applicatifs typés où l'opération d'application d'un opérateur à un opérande est considérée comme primitive. Les niveaux deux et trois s'expriment dans le formalisme de la logique combinatoire typée de H.B. Curry (1958). Cette logique fait appel à des opérateurs abstraits — appelés «combinateurs» — qui permettent de composer intrinsèquement des opérateurs plus élémentaires entre eux (Desclés 1990). Les combinateurs sont associés à des règles d'introduction et d'élimination. Ceux que nous utilisons¹ sont **B**, **C***, **S**, **Φ**, avec les règles suivantes (U₁, U₂, U₃ sont des expressions applicatives typées) :

introduction

U₁ (U₂ U₃)

----- (i-**B**)

B U₁ U₂ U₃

élimination

B U₁ U₂ U₃

----- (e-**B**)

U₁ (U₂ U₃)

¹ Le combinateur **C*** est souvent noté **T**.

$$\begin{array}{r}
 U_1 U_2 \\
 \hline
 \text{C}_* U_2 U_1
 \end{array}
 \text{---(i-C}_*\text{)}
 \qquad
 \begin{array}{r}
 \text{C}_* U_2 U_1 \\
 \hline
 U_1 U_2
 \end{array}
 \text{---(e-C}_*\text{)}$$

$$\begin{array}{r}
 U_1 (U_2 U_4) (U_3 U_4) \\
 \hline
 \Phi U_1 U_2 U_3 U_4
 \end{array}
 \text{---(i-}\Phi\text{)}
 \qquad
 \begin{array}{r}
 \Phi U_1 U_2 U_3 U_4 \\
 \hline
 U_1 (U_2 U_4) (U_3 U_4)
 \end{array}
 \text{---(e-}\Phi\text{)}$$

Ces règles conduisent à une autre formulation sous forme de règle de réduction (β -réduction) ou d'expansion (β -expansion) :

$$\begin{aligned}
 & ((\mathbf{B} U_1 U_2) U_3) \geq (U_1 (U_2 U_3)) \\
 & ((\mathbf{C}_* U_1) U_2) \geq (U_2 U_1) \\
 & ((\Phi U_1 U_2 U_3) U_4) \geq (U_1 (U_2 U_4)(U_3 U_4))
 \end{aligned}$$

Dans ce qui suit, nous nous intéresserons aux relations entre les deux premiers niveaux (le phénotype et le génotype) en mettant en oeuvre un système d'analyse formelle — appelée : Grammaire Catégorielle Combinatoire Applicative (GCCA) — qui relie explicitement les expressions phénotypiques à leurs représentations sous-jacentes dans le génotype². Ce système consiste en :

- (i) l'analyse syntaxique des expressions concaténées du phénotype par une Grammaire Catégorielle Combinatoire;
- (ii) la construction à partir du résultat de l'analyse syntaxique d'une interprétation sémantique fonctionnelle des expressions phénotypiques.

1.1 Les Grammaires Catégorielles

Des chercheurs par leurs travaux ont marqué l'évolution des Grammaires Catégorielles :

- (i) Husserl (1913) : Catégories de significations.
- (ii) Lesniewski (1922) : Catégories sémantiques.
- (iii) Ajdukiewicz (1935), Bar-Hillel (1953) : Grammaires Catégorielles.
- (iv) Lambek (1958, 1961) : Calcul de Lambek.
- (v) Steedman (1989) : Grammaire Catégorielle Combinatoire.

Les Grammaires Catégorielles assignent des catégories syntaxiques à chaque unité linguistique. Les catégories syntaxiques sont des types orientés engendrés à partir de types de base et de deux opérateurs constructifs '/' et '\'.

- (i) N (syntagme nominal) et S (phrase) sont des types de base.
- (ii) Si X et Y sont des types orientés alors X/Y et X\Y sont des types orientés³.

² Dans le phénotype, les expressions linguistiques sont concaténées selon les règles syntagmatiques du français. Dans le génotype, les expressions sont agencées selon l'ordre applicatif.

³ Nous choisissons ici la notation de Steedman (1989) : X/Y et X\Y sont des types orientés fonctionnels. Une unité linguistique 'u' avec le type X/Y (respectivement X\Y) est considérée comme un opérateur (ou une fonction) dont l'opérande de type Y est positionné à droite (respectivement à gauche) de l'opérateur.

Une unité linguistique u de type orienté X sera désigné par ' $[X : u]$ '.

Les deux règles d'application (avant et arrière) sont notées :

$$\begin{array}{ccc} [X/Y : u_1] - [Y : u_2] & & [Y : u_1] - [X \setminus Y : u_2] \\ \text{-----}> & ; & \text{-----}< \\ [X : (u_1 u_2)] & & [X : (u_2 u_1)] \end{array}$$

Les prémisses dans chaque règle sont des concaténations d'unités linguistiques à types orientés considérées comme étant des opérateurs ou des opérands, la conséquence de chaque règle est une expression applicative avec un type orienté.

La Grammaire Catégorielle Combinatoire (Steedman 1989) généralise les Grammaires Catégorielles classiques en introduisant des opérations de changement de type et des opérations de composition des types fonctionnels. Les nouvelles règles proposées visent une analyse pseudo-incrémentale (de gauche à droite) pour éliminer le problème de la pseudo-ambiguïté (Haddock 1987; Pareschi, Steedman 1987).

1.2 La Grammaire Catégorielle Combinatoire Applicative

Nous considérons dans la GCCA que les règles de la Grammaire Catégorielle Combinatoire de Steedman introduisent les combinateurs **B**, **C***, **S** dans la séquence syntagmatique. Cette introduction permet de passer d'une structure concaténée à une structure applicative.

Les règles de la GCCA sont :

Règles d'application :

$$\begin{array}{ccc} [X/Y : u_1] - [Y : u_2] & & [Y : u_1] - [X \setminus Y : u_2] \\ \text{-----}> & ; & \text{-----}< \\ [X : (u_1 u_2)] & & [X : (u_2 u_1)] \end{array}$$

Règles de changement de type :

$$\begin{array}{ccc} [X : u] & & [X : u] \\ \text{-----}>\mathbf{T} & ; & \text{-----}<\mathbf{T} \\ [Y/(Y \setminus X) : (\mathbf{C}_* u)] & & [Y \setminus (Y/X) : (\mathbf{C}_* u)] \end{array}$$

$$\begin{array}{ccc} [X : u] & & [X : u] \\ \text{-----}>\mathbf{T}x & ; & \text{-----}<\mathbf{T}x \\ [Y/(Y/X) : (\mathbf{C}_* u)] & & [Y \setminus (Y \setminus X) : (\mathbf{C}_* u)] \end{array}$$

Règles de composition fonctionnelle :

$$\begin{array}{ccc} [X/Y : u_1] - [Y/Z : u_2] & & [Y \setminus Z : u_1] - [X \setminus Y : u_2] \\ \text{-----}>\mathbf{B} & ; & \text{-----}<\mathbf{B} \\ [X/Z : (\mathbf{B} u_1 u_2)] & & [X \setminus Z : (\mathbf{B} u_2 u_1)] \end{array}$$

$$\frac{[X/Y : u_1]-[Y\backslash Z : u_2]}{[X\backslash Z : (\mathbf{B} u_1 u_2)]} \rightarrow \mathbf{Bx} \quad ; \quad \frac{[Y/Z : u_1]-[X\backslash Y : u_2]}{[X/Z : (\mathbf{B} u_2 u_1)]} \leftarrow \mathbf{Bx}$$

Les prémisses des règles sont des expressions concaténées typées; les résultats sont des expressions applicatives (typées) avec éventuellement introduction d'un combinateur. Le changement de type d'une unité u introduit le combinateur \mathbf{C}_* ; la composition de deux unités concaténées introduit le combinateur \mathbf{B} et \mathbf{S} . Avec ces règles nous pouvons analyser une phrase au moyen d'une stratégie quasi-incrémentale «de gauche à droite». Le choix d'une telle stratégie est motivé par le contrôle du problème de la pseudo-ambiguïté (Pareschi, Steedman 1987; Steedman 1989).

Exemple :

$$\begin{array}{l} \text{Jean} \quad \quad \quad - \quad \quad \quad \text{aime} \\ \text{-----} \quad \quad \quad \quad \quad \quad \text{-----} \\ [N : \text{Jean}] \\ \text{-----} \rightarrow \mathbf{T} \\ [S/(S\backslash N) : (\mathbf{C}_* \text{Jean})] \quad \quad \quad - \quad \quad \quad [(S\backslash N)/N : \text{aime}] \\ \text{-----} \rightarrow \mathbf{B} \\ [S/N : (\mathbf{B} (\mathbf{C}_* \text{Jean}) \text{aime})] \end{array}$$

La première règle (\mathbf{T}) appliquée à l'unité typée $[N : \text{Jean}]$ transforme l'opérande en opérateur. Elle construit une structure applicative $(\mathbf{C}_* \text{Jean})$ ayant pour type $S/(S\backslash N)$. L'introduction du combinateur \mathbf{C}_* illustre dans la représentation applicative le changement de type : $(\mathbf{C}_* \text{Jean})$ fonctionne comme un opérateur avec son type fonctionnel. La règle (\mathbf{B}) combine les unités linguistiques typées $[S/(S\backslash N) : (\mathbf{C}_* \text{Jean})]$ et $[(S\backslash N)/N : \text{aime}]$ avec le combinateur \mathbf{B} de façon à pouvoir composer les deux unités fonctionnelles $(\mathbf{C}_* \text{Jean})$ et aime .

Un traitement complet basé sur la Grammaire Catégorielle Combinatoire Applicative s'effectue en deux grandes étapes :

- (i) la première étape s'illustre par la vérification de la bonne connexion syntaxique et la construction de structures prédicatives avec des combinateurs introduits à certaines positions de la chaîne syntagmatique;
- (ii) la deuxième étape consiste à utiliser les règles de β -réduction des combinateurs de façon à former une structure prédicative sous-jacente à l'expression phénotypique. L'expression obtenue est applicative et appartient au langage génotype. La GCCA engendre des processus qui associent une structure applicative à une expression concaténée du phénotype. Il nous reste à éliminer les combinateurs de l'expression obtenue de façon à construire la «forme normale» (au sens technique de la β -réduction) qui exprime l'interprétation sémantique fonctionnelle. Ce calcul s'effectue entièrement dans le génotype.

Le traitement que nous proposons donc prend la forme d'une compilation dont les étapes sont résumées dans la figure 1 :

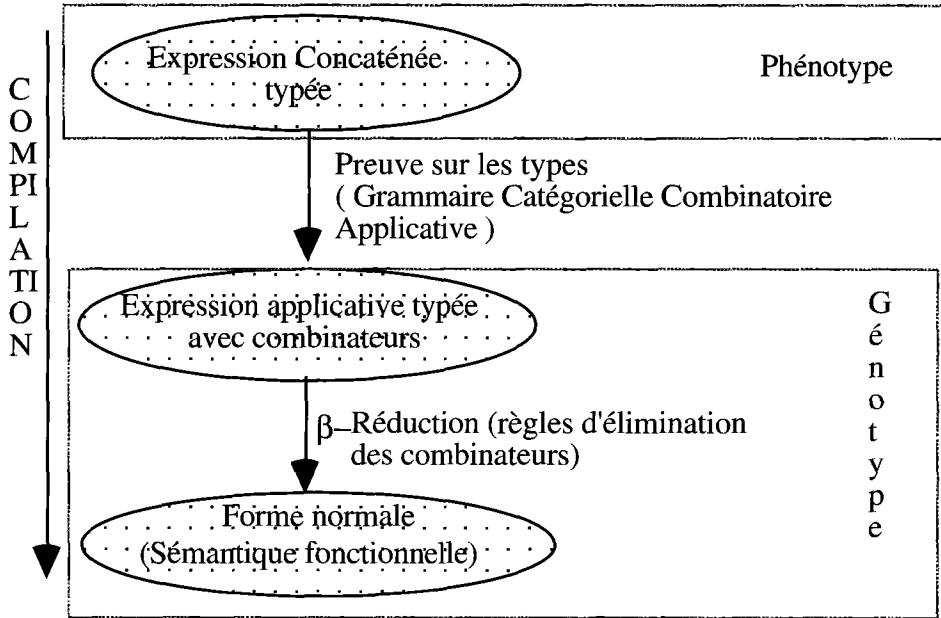


Figure 1

Traitons un exemple simple : *Jean aime Marie*

1 [N :Jean]-[(S\N)/N :aime]- [N :Marie]		Structure concaténée typée du phénotype
2 [S/(S\N) : (C _* Jean)]-[(S\N)/N :aime]- [N :Marie]	(>T)	Compilation
3 [S/N : (B (C _* Jean) aime)]-[N :Marie]	(>B)	
4 [S : ((B (C _* Jean) aime) Marie)]	(>)	
	∇	
5 [S : ((B (C _* Jean) aime) Marie)]		Structure applicative typée du génotype
6 [S : ((C _* Jean) (aime Marie))] (B)		
7 [S : ((aime Marie) Jean)] (C _*)		Forme normale du génotype

Le changement de type (>T) affectant l'opérande *Jean* permet d'engendrer l'opérateur (C_{*} Jean) que la règle fonctionnelle (>B) compose avec l'opérateur *aime*. L'opérateur complexe (B (C_{*} Jean) aime) s'applique à l'opérande *Marie* pour former

l'expression applicative du génotype ((B (C_{*} Jean) aime) Marie). La réduction des combinateurs dans le génotype construit l'interprétation sémantique fonctionnelle sous-jacente à l'expression phénotypique en entrée.

Les règles catégorielles combinatoires applicatives ne sont pas conçues pour réaliser un simple calcul. Un principe purement applicatif motive la conception de telles règles. Rappelons qu'avec les Grammaires Catégorielles, les unités linguistiques sont considérées comme des opérateurs avec leurs opérands. De ce fait, avec les règles d'applications fonctionnelles nous formalisons l'application d'un opérateur à son opérande, avec les règles de composition nous composons deux foncteurs pour construire un foncteur complexe.

2. LA RÉORGANISATION STRUCTURELLE

L'analyse syntaxique «de gauche à droite» soulève le problème du non-déterminisme introduit par la présence dans la langue, de modifieurs arrières qui sont des opérateurs qui s'appliquent à l'ensemble ou à une partie d'une structure préalablement construite.

Si dans le premier cas l'utilisation d'une règle d'application permet la poursuite de l'analyse⁴, il en est autrement pour le second cas où l'analyse «patine». Pour une phrase comme *Jean aime Marie tendrement* l'analyseur produit dans un premier temps le constituant [S :((B(C_{*} Jean) aime) Marie)]. Ce dernier n'est pas combinable avec *tendrement*, de type (S\N)\(S\N). En effet, *tendrement* est un opérateur qui a comme opérande (*aime Marie*) positionné à sa gauche. Une analyse quasi-incrémentale de «gauche à droite» favorise l'application d'une règle combinatoire dès que possible. Ce facteur a pour conséquence directe de «noyer»⁵ *aime* et *Marie* dans ((B(C_{*} Jean) aime) Marie), ce qui évidemment ne nous permet pas de construire directement l'opérande (*aime Marie*).

Le problème posé revient à la possibilité d'un retour arrière. Mais ce retour arrière est de nature à accroître le coût «computationnel» (mémoire et temps d'exécution) d'une analyse syntaxique. Cependant, un retour arrière «intelligent» (que nous allons proposer plus loin) peut nous permettre de réduire considérablement ce coût tout en construisant des analyses sémantiques correctes et en éliminant les pseudo-ambiguïtés.

Ainsi un tel retour arrière décomposera le constituant déjà construit en deux composantes dont une se combine forcément avec le modifieur arrière.

Formellement, cette opération de réorganisation structurelle s'effectue par les deux étapes successives suivantes :

4 Prenons l'exemple de la phrase *Jean frappa Marie hier* où le modifieur arrière *hier* opère sur l'ensemble de la phrase *Jean frappa Marie*; *hier* étant de type syntaxique S\S, il suffit, pour poursuivre l'analyse, d'appliquer *hier* à *Jean frappa Marie* par la règle (<).

5 C'est à dire que *Marie* n'apparaît pas comme l'opérande directe de l'opérateur *aimer*.

a- La réorganisation du constituant déjà construit isole à chaque fois deux sous-catégories, et teste si le modifieur arrière se combine à gauche⁶, ou pas, avec une de ces deux sous-catégories. Nous procédons ensuite à la réduction des combinateurs jusqu'à ce que le test nous donne une valeur positive. À la fin du processus nous récupérons une nouvelle structure applicative typée «équivalente» à la première.

Exemple : Dans le cas de l'énoncé *Jean aime Marie tendrement*, les étapes de la réorganisation sont :

Le constituant construit : [S : ((**B** (**C**_{*} *Jean*) *aime*) *Marie*)]

Les deux sous-catégories sont : [S/N : (**B** (**C**_{*} *Jean*) *aime*)]; [N : *Marie*]

Test : [S/N : (**B** (**C**_{*} *Jean*) *aime*)] ne se combine pas à gauche avec

[(S\N)\(S\N) : *tendrement*]

[N : *Marie*] ne se combine pas à gauche avec

[(S\N)\(S\N) : *tendrement*]

Réduction du combinateur **B** : [S : ((**C**_{*} *Jean*) (*aime Marie*))]

Les deux sous-catégories sont : [S/(S\N) : (**C**_{*} *Jean*)];

[S\N : (*aime Marie*)]

Test : [S/(S\N) : (**C**_{*} *Jean*)] ne se combine pas à gauche avec

[(S\N)\(S\N) : *tendrement*]

[S\N : (*aime Marie*)] se combine à gauche avec

[(S\N)\(S\N) : *tendrement*]

Arrêt du processus de réduction des combinateurs. Nous récupérons en sortie la catégorie :

[S : ((**C**_{*} *Jean*) (*aime Marie*))].

b- La décomposition réalisée grâce aux deux règles⁷ :

[X : (u₁ u₂)]

[X : (u₁ u₂)]

----->**dec**

;

-----<**dec**

[X/Y : u₁]-[Y : u₂]

[Y : u₂]-[X\Y : u₁]

Nous lisons ces règles comme suit :

6 Dans notre terminologie, u₁ se combine à gauche avec u₂ si une des règles suivantes <, <**B**, <**Bx**, <**S**, <**Sx**, peut composer les types de u₂ et u₁ ou si une des règles combinatoires peut composer les types de (**C**_{*} u₂) et u₁.

7 Notons que les deux règles (>**dec**) et (<**dec**) sont respectivement inverses des règles d'application fonctionnelle (>) et (<).

- Pour (**>dec**) : Si nous avons une structure applicative ($u_1 u_2$) de type X, avec u_1 de type X/Y et u_2 de type Y, alors nous pouvons construire une nouvelle expression concaténée formée des deux catégories $[X/Y:u_1]$ et $[Y:u_2]$.

- Pour (**<dec**) : Si nous avons une structure applicative ($u_1 u_2$) de type X, avec u_1 de type X\Y et u_2 de type Y, alors nous pouvons construire une nouvelle expression concaténée formée des deux catégories $[Y:u_2]$ et $[X\Y:u_1]$.

Ces deux règles nous permettent de reconstruire un nouvel agencement concaténé de la structure opérateur/opérande issue de la réorganisation.

Pour la phrase *Jean aime Marie tendrement* la décomposition est appliquée à la structure qui résulte de la réorganisation :

$$[S : ((C_* Jean) (aime Marie))]$$

Avec la règle (**>dec**), nous produisons l'agencement concaténé :

$$[S/(S\N) : (C_* Jean)] - [S\N : (aime Marie)].$$

Ces deux étapes entrent dans l'analyse complète de la phrase *Jean aime Marie tendrement* comme suit (étape 5 pour la réorganisation et étape 6 pour la décomposition) :

Structure concaténée typée du phénotype

1	$[N : Jean] - [(S\N)/N : aime] - [N : Marie] - [(S\N)\(S\N) : tendrement]$	
...		
4	$[S : ((B (C_* Jean) aime) Marie)] - [(S\N)\(S\N) : tendrement]$	
5	$[S : ((C_* Jean) (aime Marie))] - [(S\N)\(S\N) : tendrement]$	(a)
6	$[S/(S\N) : (C_* Jean)] - [S\N : (aime Marie)] - [(S\N)\(S\N) : tendrement]$	(b)
7	$[S/(S\N) : (C_* Jean)] - [S\N : (tendrement (aime Marie))]$	(<)
8	$[S : ((C_* Jean) (tendrement (aime Marie)))]$	(>)

Structure applicative typée du génotype

9	$[S : ((C_* Jean) (tendrement (aime Marie)))]$	
10	$[S : ((tendrement (aime Marie)) Jean)]$	(C_*) Forme normale du génotype

3. LA COORDINATION

La coordination est l'action de joindre deux mots ou deux expressions du même genre ou ayant même fonction. Dans le cadre des Grammaires Catégorielles, Steedman (1989), Barry et Pickering (1990) considèrent que deux unités linguistiques peuvent être coordonnées pour donner une unité linguistique de type X si et seulement si chaque unité est de type X. Même si cette définition reste incomplète, sachant que la coordination se

présente sous différentes formes, elle indique de manière idéale la voie à suivre pour établir une solution fiable.

Présentons quatre types d'exemples de coordination par ET. Nous pouvons coordonner⁸ :

a) Deux segments de même type, de même structure et contigus à ET :

[Jean aime]S/N et [Paul déteste]S/N ces tableaux

b) Deux segments dans une construction elliptique :

Jean aime [Marie tendrement] et [Sophie sauvagement]

[Jean] aime [Marie] et [Paul Sophie]

c) Deux segments de structures différentes :

Marie marche [doucement] et [avec élégance]

Jean [chante] et [joue du violon]

d) Deux segments, sans distributivité :

Le drapeau est [blanc] et [rouge]

(≠ Le drapeau est blanc et le drapeau est rouge).

À la conjonction 'ET' on associe le type polymorphique (XX)/X. Le contexte donne cependant plus de spécifications pour assigner un type à 'ET'.

Les hypothèses 1 et 2 permettent d'assigner un type à ET en tenant compte du contexte.

Hypothèse 1 : *La catégorie construite qui suit immédiatement la conjonction ET détermine le type de la coordination.*

Cette hypothèse nous amène indirectement à introduire une interruption dans l'analyse quasi-incrémentale : dès que nous rencontrons la conjonction ET nous interrompons momentanément l'analyse quasi-incrémentale pour construire le second membre de la coordination.

Nous revenons ensuite en arrière pour déterminer le premier membre de la coordination.

Nous proposons la seconde hypothèse :

Hypothèse 2 : *Quand nous avons une coordination de type X définie par l'hypothèse 1, le premier membre de la coordination est la catégorie de type X qui précède immédiatement la conjonction.*

Les règles que nous devons dégager à travers ces deux hypothèses partent donc de l'idée que les deux membres de la coordination ont des types syntaxiques X identiques correspondant à des interprétations sémantiques fonctionnelles différentes. Le résultat de

⁸ Les catégories à coordonner sont entre crochets.

l'application de ces règles conserve le même type syntaxique X. Nous établissons deux types abstraits pour la conjonction. Le premier concerne la conjonction distributive, nous le noterons CONJD. Le deuxième type concerne la conjonction non distributive et nous le noterons CONJN.

[X : u ₁]-[CONJD : et]-[X : u ₂]	[X : u ₁]-[CONJN : et]-[X : u ₂]
<CONJD>;	<CONJN>
[X : (Φ et u ₁ u ₂)]	[X : (et u ₁ u ₂)]

Nous appliquons la règle <CONJD> aux cas de la coordination distributive. Pour prendre en compte la distributivité au niveau de la structure applicative, nous utilisons le combinateur Φ. Nous appliquons la règle <CONJN> aux cas de la coordination non distributive (voir exemple E3).

Avec l'analyse quasi-incrémentale, lors de l'application de l'hypothèse 2, deux cas de figure se présentent :

- Le constituant produit avant de rencontrer la conjonction est du même type que le constituant déterminé par la coordination. Ce constituant est alors le premier membre de la coordination. Par exemple, l'analyse de la phrase : [*Jean aime*]_{S/N} et [*Paul détecte*]_{S/N} ces tableaux construit [S/N : (B(C_{*}Jean) aime)] avant de rencontrer la conjonction. Le type est identique au type du second membre [S/N : (B(C_{*}Paul) détecte)], le constituant déterminé par la première hypothèse. Le constituant [S/N : (B(C_{*}Jean) aime)] est alors le premier membre de la coordination.

- Le constituant déterminé avant de rencontrer la conjonction n'est pas du même type que le constituant déterminé par la coordination. Il est nécessaire de modifier la structure de ce constituant. Par exemple, l'analyse de la phrase : *Jean aime* [*Marie tendrement*] et [*Sophie sauvagement*] construit [S : ((C_{*} Jean) (tendrement (aime Marie)))] avant l'analyse de la conjonction. Le second membre de la coordination étant [(S/N)\((S/N)/N) : (B sauvagement (C_{*} Sophie))]⁹.

Dans ce second cas le processus de réorganisation structurelle nous permet :

- soit d'isoler directement le premier membre de la coordination¹⁰.
- soit d'isoler la structure binaire «opérateur/opérande» qui contient le premier membre de la coordination. Dans cette deuxième situation, il est nécessaire d'adjoindre à la réorganisation structurelle l'utilisation des équivalences logiques¹¹ de la logique combinatoire (e,f,g,h)¹² :

$$(e) \quad (u_1 (u_2 u_3)) \quad \iff \quad ((B u_1 u_2) u_3)$$

9 L'énoncé *Jean aime Marie tendrement et Sophie sauvagement* est ambigu. Dans notre exemple, nous considérons que *Sophie* est objet.

10 Voir les étapes 6 et 7 de l'exemple E1.

11 Ces équivalences sont les conséquences directes de l'introduction et l'élimination des combinateurs B et C_{*}.

12 Voir l'étape 8 de l'exemple E2.

- | | | | |
|-----|--------------|-----|--|
| (f) | ((u1 u2) u3) | <=> | ((B (C _* u3) u1) u2) |
| (g) | (u1 (u2 u3)) | <=> | ((B u1 (C _* u3)) u2) |
| (h) | ((u1 u2) u3) | <=> | ((B (C _* u3) (C _* u2)) u1) |

4. LES MÉTARÈGLES

Nous enrichissons notre formalisme par différentes métarègles qui contrôlent le changement de type.

Ces métarègles d'une part nous indiquent qu'une règle de changement de type doit être appliquée, et d'autre part choisissent le changement de type particulier à effectuer.

Nous ne concevons pas ces métarègles comme un outil purement informatique, nous leur donnons une pertinence logique et linguistique. Elles peuvent avoir une interprétation si nous tenons compte des facteurs prosodiques.

Nous présentons dans ce qui suit trois métarègles parmi celles que nous avons conçues. Ces dernières sont au nombre de dix¹³.

Soient u₁ et u₂ dans l'expression concaténée 'u₁-u₂' :

Métarègle 1 : Si u₁ est de type N et u₂ de type (Y\N)/Z, alors nous appliquons le changement de type avant (>T) à u₁ :

[N : u₁ ==> Y/(Y\N) : (C_{*} u₁)]

Exemple : *Jean mange la pomme*

[N : *Jean*]-[(S\N)/N : *mange*]-[N/N : *la*]-[N : *pomme*]

[S/(S\N) : (C_{*} *Jean*)]-[(S\N)/N : *mange*]-[N/N : *la*]-[N : *pomme*]

Dans ce cas Y = S; Z = N.

Métarègle 2 : Si u₁ est de type N (u₁ précédé de *et*) et u₂ de type N, alors nous appliquons le changement de type avant (>T) à u₁ :

[N : u₁ ==> S/(S\N) : (C_{*} u₁)]

Exemple : *Jean aime Marie et Paul Sophie*

...-[CONJD : *et*]-[N : *Paul*]-[N : *Sophie*]

...-[CONJD : *et*]-[S/(S\N) : (C_{*} *Paul*)]-[N : *Sophie*]

Métarègle 3 : Si u₂ est de type N et u₁ de type Y/X (u₁ précédé de *et*), alors nous appliquons le changement de type arrière (<T) à u₂ :

[N : u₂ ==> X/(X/N) : (C_{*} u₂)]

Exemple : *Jean aime Marie et Paul Sophie*

¹³ Voir Biskri (1995).

...-[CONJD : et]-[S/(S\N) : (C* Paul)]-[N : Sophie]

...-[CONJD : et]-[S/(S\N) : (C* Paul)]-[(S\N)\((S\N)/N) : (C* Sophie)]

Dans ce cas X = S\N; Y = S.

5. EXEMPLES

E1 : Jean aime Marie et déteste Sophie

Phénotype

1	[N:Jean]-[(S\N)/N:aime]-[N:Marie]-[CONJD:et]-[(S\N)/N:déteste]-[N:Sophie]	
4	... [S:((B (C* Jean) aime) Marie)]-[CONJD:et]-[(S\N)/N:déteste]-[N:Sophie]	
5	[S:((B (C* Jean) aime) Marie)]-[CONJD:et]-[S\N:(déteste Sophie)]	(>)
6	[S:((C* Jean) (aime Marie))]-[CONJD:et]-[S\N:(déteste Sophie)]	(B)
7	[S/(S\N):(C* Jean)]-[S\N:(aime Marie)]-[CONJD:et]-[S\N:(déteste Sophie)]	(>dec)
8	[S/(S\N):(C* Jean)]-[S\N:(Φ et (aime Marie) (déteste Sophie))]	(<CONJD>)
9	[S:((C* Jean) (Φ et (aime Marie) (déteste Sophie)))]	(>)

Génotype

10	[S : ((C* Jean) (Φ et (aime Marie) (déteste Sophie)))]	
11	[S : ((Φ et (aime Marie) (déteste Sophie)) Jean)]	(C*)
12	[S : (et ((aime Marie) Jean) ((déteste Sophie) Jean))]	(Φ)

E2 : Jean aime Marie et Paul Sophie

Phénotype

1	[N:Jean]-[(S\N)/N:aime]-[N:Marie]-[CONJD:et]-[N:Paul]-[N:Sophie]	
4	... [S:((B (C* Jean) aime) Marie)]-[CONJD:et]-[N:Paul]-[N:Sophie]	
5	...-[CONJD:et]-[S/(S\N):(C* Paul)]-[N:Sophie]	(>T), M2
6	...-[CONJD:et]-[S/(S\N):(C* Paul)]-[(S\N)\(S/(S\N)):(C* Sophie)]	(<T), M3
7	...-[CONJD:et]-[S\N(S/(S\N)):(B (C* Paul)(C* Sophie))]	(>Bx)
8	[S:((B (C* Jean) (C* Marie) aime)]-[CONJD:et]-...	(h)
9	[(S\N)/N:aime]-[S\N(S/(S\N)):(B (C* Jean) (C* Marie))]-[CONJD:et]-...	(<dec)
10	[-[S\N(S/(S\N)):(Φ et (B(C* Jean)(C* Marie))(B(C* Paul)(C* Sophie)))]	(<CONJD>)
11	[S:((Φ et (B (C* Jean) (C* Marie))(B (C* Paul)(C* Sophie))) aime]	(<)

Génotype

12	[S : ((Φ et (B (C _* Jean) (C _* Marie))(B (C _* Paul)(C _* Sophie))) aime)]	
13	[S : (et ((B (C _* Jean) (C _* Marie)) aime) ((B (C _* Paul)(C _* Sophie)) aime))] (Φ)	
14	[S : (et ((C _* Jean) ((C _* Marie) aime)) ((B (C _* Paul)(C _* Sophie)) aime))] (B)	
15	[S : (et (((C _* Marie) aime) Jean) ((B (C _* Paul)(C _* Sophie)) aime))] (C _*)	
16	[S : (et ((aime Marie) Jean) ((B (C _* Paul)(C _* Sophie)) aime))] (C _*)	
17	[S : (et ((aime Marie) Jean) ((C _* Paul)((C _* Sophie) aime))] (B)	
18	[S : (et ((aime Marie) Jean) (((C _* Sophie) aime) Paul))] (C _*)	
19	[S : (et ((aime Marie) Jean) ((aime Sophie) Paul))] (C _*)	

E3 : le drapeau est blanc et rouge

\Rightarrow le drapeau est blanc et le drapeau est rouge

Phénotype

1	[N/N:le]-[N:drapeau]-[(S\N)/(N\N):est]-[N\N:blanc]-[CONJN:et]-[N\N:rouge]	
2	[N:(le drapeau)]-[(S\N)/(N\N):est]-[N\N:blanc]-[CONJN:et]-[N\N:rouge]	
3	[S/(S\N):(C _* (le drapeau))]-[(S\N)/(N\N):est]-[N\N:blanc]-[CONJN:et]-... (>T), M1	
4	[S/(N\N):(B (C _* (le drapeau)) est)]-[N\N:blanc]-[CONJN:et]-[N\N:rouge] (>B)	
5	[S:(B (C _* (le drapeau)) est) blanc]-[CONJN:et]-[N\N:rouge] (>)	
6	[S/(N\N):(B (C _* (le drapeau)) est)]-[N\N:blanc]-[CONJN:et]-[N\N:rouge] (>dec)	
7	[S/(N\N):(B (C _* (le drapeau)) est)]-[N\N:(et blanc rouge)] (<CONJN>)	
8	[S:(B (C _* (le drapeau)) est) (et blanc rouge)] (>)	

Génotype

9	[S : ((B (C _* (le drapeau)) est) (et blanc rouge)))]	
10	[S : ((C _* (le drapeau)) (est (et blanc rouge)))] (B)	
11	[S : ((est (et blanc rouge)) (le drapeau))] (C _*)	

D'autres exemples et plus de détails sont donnés dans (Biskri 1995). Les analyses sont implémentées. Nous ne donnons pas ici les détails de l'algorithme.

6. CONCLUSION

Avec ce travail, nous sommes arrivés à la conviction que les Grammaires Catégorielles associées à un formalisme applicatif peuvent rendre compte correctement des langues naturelles. Même si, beaucoup de choses restent à faire pour établir un système prêt à répondre à des besoins industriels, nous pensons que ce que nous présentons est

assez solide pour être considéré comme un modèle de traitement des langues naturelles relativement fiable.

Ce modèle, d'ailleurs, présente des qualités que nous jugeons importantes. En premier lieu, nous trouvons qu'il est élégant dans le sens où il permet un suivi de traitement très aisé. Deuxièmement, l'utilisation combinée des types syntaxiques et des combinateurs permet de véhiculer un maximum d'informations avec un minimum de ressources (variables). Cette propriété, même si nous sommes en face uniquement d'une étude théorique, a l'avantage de réduire (voire éliminer) des traitements que nous dirons superflus, comme celui fondé sur l'unification pour retrouver les interprétations sémantiques dans le modèle de Steedman et Pareschi (1987).

Par ailleurs, le modèle de la GCCA prend la forme d'une compilation. On ne fait pas un traitement sémantique parallèle à un traitement syntaxique. La construction de l'interprétation sémantique fonctionnelle se fait dans la continuité du calcul syntaxique avec le passage du phénotype au génotype.

Comme nous l'avons signalé précédemment, ce travail s'inscrit dans un cadre de travail beaucoup plus général. D'autres traitements dans le génotype suivent celui que nous avons présenté. Nous pensons en particulier au traitement des paraphrases, de la passivation et de la reflexivation etc. Ainsi, nous pouvons conseiller au lecteur de consulter (Desclés 1990) pour comprendre comment peuvent se faire ces genres de traitements.

Pour résumer, ce modèle est capable de réaliser pour un noyau du français, les objectifs suivants :

- (i) produire une analyse qui vérifie la correction syntaxique des énoncés;
- (ii) engendrer automatiquement des structures prédicatives qui rendent compte de l'interprétation sémantique fonctionnelle des énoncés.

Ce modèle a en outre les propriétés suivantes :

- 1- Un premier calcul vérifie la correction syntaxique, ce calcul se poursuit par une construction de l'interprétation sémantique fonctionnelle. Cela est rendu possible par l'introduction des combinateurs à des positions spécifiques de la chaîne syntagmatique.
- 2- Nous introduisons des éléments de nature sémantique fonctionnelle par des outils syntaxiques applicatifs (les combinateurs).
- 3- Nous calculons l'interprétation sémantique fonctionnelle par des méthodes syntaxiques applicatifs (la réduction des combinateurs).

Pour résumer, nous interprétons à l'aide de techniques purement syntaxiques. La distinction syntaxe/sémantique devrait alors être repensée dans une autre perspective.

RÉFÉRENCES

- ADES, A., STEEDMAN, M. (1982) : «On the order of words», *Linguistics and Philosophy*, 4, pp. 517-558.
- BARRY, G., PICKERING, M. (1990) : «Dependancy and constituency in categorial grammar», dans Lecomte 1992, *L'ordre des mots dans les grammaires catégorielles*, pp. 38-57.
- BISKRI, I. (1995) : *La Grammaire Catégorielle Combinatoire Applicative dans le cadre de la Grammaire Applicative et Cognitive*, Thèse de Doctorat, EHESS, Paris.
- BISKRI, I., MEUNIER, J.G., NAULT, G. (1997) : «Extraction des connaissances terminologiques au moyen des Grammaires Catégorielles : un modèle hybride», Actes du colloque JST97, Avignon, France.
- CURRY, B.H., FEYS, R. (1958) : *Combinatory logic*, Vol. I, North-Holland.
- DESCLÉS, J.P. (1990) : *Langages applicatifs, langues naturelles et cognition*, Paris, Hermes.
- DESCLÉS, J.P., SEGOND, F. (1990) : «Topicalization: categorial analysis and applicative grammar», dans Lecomte 1992, *L'ordre des mots dans les grammaires catégorielles*, pp. 13-37.
- DESCLÉS, J.P., BISKRI, I. (1996) : «Logique combinatoire et linguistique : Grammaire Catégorielle Combinatoire Applicative», *Revue Mathématiques, Informatiques et Sciences Humaines*, Paris.
- HADDOCK, N. (1987) : «Incremental interpretation and Combinatory Categorial Grammar», dans N. Haddock, E. Klein, et G. Morill, *Working papers in cognitive science volume I: Categorial Grammar, Unification Grammar and Parsing*, Edinburgh Univ., pp. 71-84.
- LECOMTE, A. (1994) : *Modèles Logiques en Théorie Linguistique : éléments pour une théorie informationnelle du langage*, Synthèse de travaux présentés en vue de l'habilitation à diriger des recherches.
- MOORTGAT, M. (1988) : *Categorial Investigation, Logical and Linguistic Aspects of the Lambek Calculus*, Foris Publications.
- OEHRLE, R.T., BACH, E., WHEELER, D. (1988) : *Categorial Grammars and Natural Languages Structures*, D. Reidel.
- PARESCHI, R., STEEDMAN, M. (1987) : «A lazy Way to chart parse with categorial grammars», *Acte du colloque ACL*, Stanford.
- SHAUMYAN, S.K. (1987) : *A Semiotic Theory of Natural Language*, Bloomington, Indiana Univ. Press.
- STEEDMAN, M. (1989) : *Work in progress: Combinators and grammars in natural language understanding*, Summer Institute of Linguistic, Tucson University.
- SZABOLCSI, A. (1987) : «On combinatory categorial grammar», *Actes du Symposium on logic and languages*, Budapest, Debrecen, Akademiai Kiado, pp. 151-162.

LA FORME DES MOTS NOUS RENSEIGNE-T-ELLE SUR LEUR NATURE ?

Sylviane CARDEY⁽¹⁾, Zahra EL HAROUCHY⁽¹⁾, Peter GREENFIELD⁽²⁾

(1) Centre de recherche Lucien Tesnière; (2) Laboratoire d'Informatique de Besançon, Besançon, France

Une des étapes du traitement automatique des langues (souvent la première) consiste à déterminer à quelle catégorie grammaticale appartient chacune des unités ou chacun des «mots» du texte traité, en d'autres termes, elle consiste à déterminer la **nature des «mots»**. Aussi, nous voudrions montrer ici comment un «mot», par ses caractéristiques morphologiques, peut nous renseigner sur sa nature. Nous présentons un dictionnaire dit en *intention* par opposition aux dictionnaires dits en *extension* souvent utilisés pour essayer de réaliser cette tâche en traitement automatique du français. Ce dictionnaire permet de gagner en temps de traitement et en espace de mémoire pour la machine. Au lieu d'avoir comme entrées, ainsi que nous les trouvons dans les dictionnaires électroniques, chacune des formes canoniques ou fléchies (mot-forme) de la langue française associée à une classe, notre dictionnaire électronique (de très petite dimension) présente des entrées sous forme de caractéristiques morphologiques classées en ensembles et sous-ensembles et ordonnées selon certaines contraintes; chacune des caractéristiques définies étant associée à une des parties du discours.

Toute analyse des langues implique, entre autres tâches, la classification des éléments de la langue traitée. Les définitions des classes de mots appelées aussi, depuis longtemps, parties du discours ou catégories grammaticales, n'ont cessé d'être discutées; nous utilisons le découpage en classes traditionnel : *le substantif, l'adjectif, le pronom, le verbe, l'article, l'adverbe, la préposition, la conjonction et l'interjection*.

LES UNITÉS TRAITÉES

Nous ne traitons, avec notre dictionnaire, que les mots simples du français. Ce qu'il faut signaler est qu'un mot simple pour une machine (puisque notre travail est réalisé pour être confié à un ordinateur) n'est pas la même chose qu'un mot simple pour un humain. Le mot simple en Traitement Automatique des Langues est une chaîne de caractères comprise entre blanc(s) et/ou signe(s) de ponctuation. Dans les exemples qui suivent : *porte, portefeuille, porte-monnaie, porte-à-faux, porte à porte, œil d'une porte, porte la main à*, seul '*porte*' est un mot simple pour nous alors que pour la machine '*portefeuille*' le sera également, voire même dans certains cas '*porte-monnaie*' et '*porte-à-faux*'. Nous traitons de ce fait les composés soudés.

SUFFIXES ET TERMINAISONS

«On peut classer les suffixes selon leur étymologie mais il est préférable, dans une grammaire décrivant la langue d'aujourd'hui, de les inventorier plutôt en rapport avec leur fonctionnement — c'est-à-dire d'après la nature des mots qu'ils servent à former; suffixes formant des noms et/ou des adjectifs (dérivation appelée souvent nominale), suffixes formant des verbes (dérivation verbale)...» (Grevisse 1988 : 246). La plupart des ouvrages de grammaire mentionnent déjà le fait que certains suffixes donnent une indication sur la catégorie à laquelle un mot appartient; certains suffixes servent à former plutôt des noms, d'autres plutôt des adverbes. Pour n'en retenir que quelques-uns, (Dubois 1995 : 10-15) :

suffixes servant à former des noms		suffixes servant à former des adjectifs	
-ace	<i>populace</i>	-able, -ble	<i>aimable, audible</i>
-ade	<i>orangeade</i>	-aire	<i>solaire, polaire</i>
-age	<i>balayage</i>	-ique	<i>chimique, ironique</i>

suffixe servant à former des adverbes :	-ment	<i>gentiment</i>
suffixes servant à former des verbes :	-ailler	<i>tournailler</i>
	-asser	<i>rêvasser</i>

Il est intéressant de voir que la grammaire normative classe aussi les suffixes d'après la nature des mots qu'ils servent à former.

Aussi, nous avons voulu tenter l'expérience d'un classement non pas uniquement des suffixes au sens traditionnel mais de ce que nous appelons terminaisons des mots; chacune de ces terminaisons devant par la suite déterminer à quelle(s) catégorie(s) grammaticale(s) les mots qui la possèdent appartiennent. Nos terminaisons ne sont qu'un certain nombre de lettres allant de 1 à 8 qui se trouvent à la fin du mot. On peut se demander s'il y a quelque rapport entre la terminaison ainsi définie et la classe, étymologiquement ou historiquement, parfois la réponse est oui mais le plus souvent, à notre connaissance, il n'y a aucun rapport.

NOTRE DICTIONNAIRE AUTOMATIQUE

La linguistique informatique a demandé et permis la création de dictionnaires à usages multiples. Il existe aujourd'hui différents dictionnaires électroniques; des dictionnaires élaborés spécialement pour le traitement automatique de la langue qu'il ne faut pas confondre avec les dictionnaires d'usage qui ne sont que des variantes des versions sur papier présentées sur support magnétique.

L'architecture d'un dictionnaire, quelle qu'elle soit, est toujours symbolisée par une entrée et une sortie. Cette architecture reflète non seulement la façon dont est conçu le dictionnaire mais également son but. Le produit final, le dictionnaire, ayant une ou plusieurs applications. La construction des dictionnaires électroniques varie : soit le dictionnaire contient l'ensemble des formes canoniques et fléchies soit il génère les différentes formes que peuvent prendre les mots à partir d'une forme canonique donnée et de règles flexionnelles. Il y a d'autres types de dictionnaires mais nous ne citons que ces derniers qui cherchent, entre autres, à atteindre le même but que nous, c'est-à-dire déterminer la ou les catégorie(s) grammaticale(s) à laquelle ou auxquelles un mot peut appartenir.

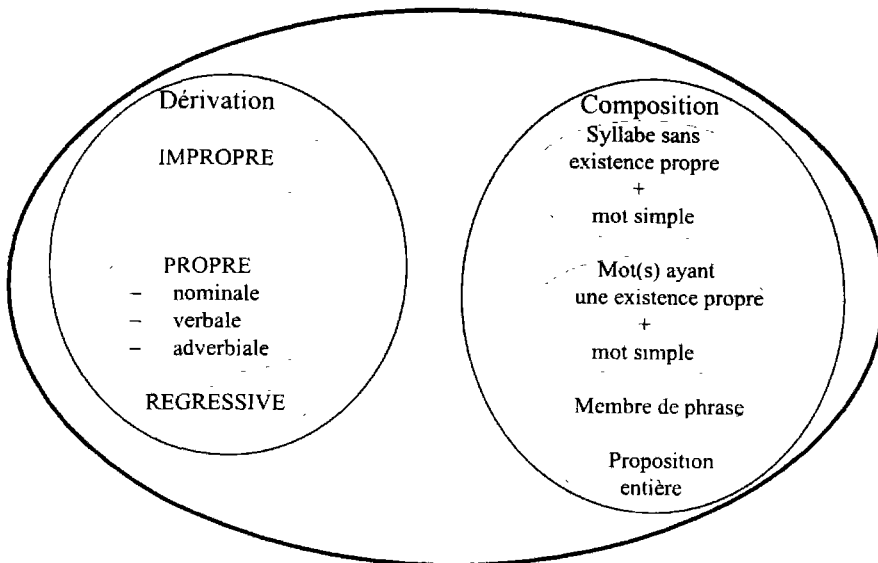
Le dictionnaire présenté ici, que nous appelons dictionnaire automatique, met en œuvre une reconnaissance de la catégorie grammaticale des mots par l'analyse de leur

terminaison. Pour construire ce dictionnaire, nous avons recensé toutes les terminaisons susceptibles d'être retenues, gardé certaines d'entre elles et rejeté les autres en fonction du nombre de «mots» concernés puis nous les avons classées. «La linguistique historique n'explique un mot *b* par un mot *a* précédent que si le passage de *a* à *b* est le cas particulier d'une règle générale valable pour bien d'autres mots, et fait comprendre ainsi que *a* soit devenu *b'*, *a''* devenu *b''*, etc. Cette régularité implique que la différence entre *a* et *b* tient à tel ou tel de leurs constituants et que dans tous les autres mots où ce constituant apparaît, il soit affecté par le même changement» (Ducrot et Schaeffer 1995 : 21). Notre dictionnaire a été créé sur ce principe. Une terminaison, pour être retenue, doit concerner un groupe homogène (mots appartenant à la même catégorie grammaticale). Nous avons relevé, lorsqu'il y en avait, les exceptions, soit les mots appartenant à des catégories grammaticales différentes de la catégorie dominante. Dans le cas où les exceptions étaient trop nombreuses, autrement dit, si les différents mots du groupe, ayant cette terminaison, appartenaient à des catégories disparates, nous n'avons pas tenu compte de la terminaison telle quelle mais avons élargi la terminaison en question à une voyelle et/ou une consonne la précédant afin de voir si ces dernières n'avaient pas une incidence sur la nature des mots.

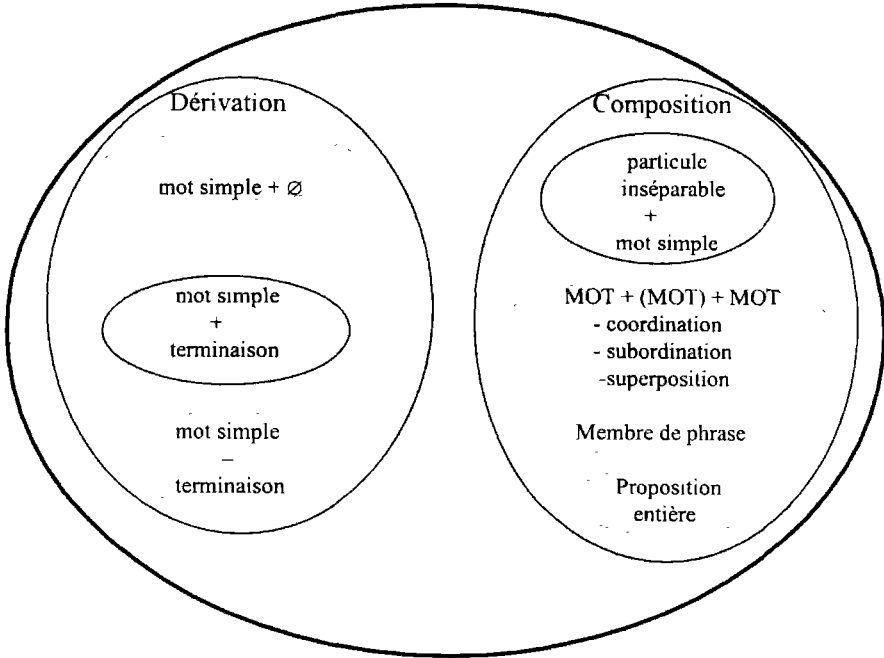
Par exemple nous avons pour la terminaison **aire** : les mots terminés par **aire** sont des *noms* sauf les mots terminés par **baire** qui sont des *adjectifs*, ceux terminés par **faire** qui sont des *verbes à l'infinitif*, ceux en **iaire** qui sont des *adjectifs* et ceux en **laire** qui sont également des *adjectifs*.

LE SYSTÈME DE LA MORPHOLOGIE LEXICALE ET LE DICTIONNAIRE AUTOMATIQUE

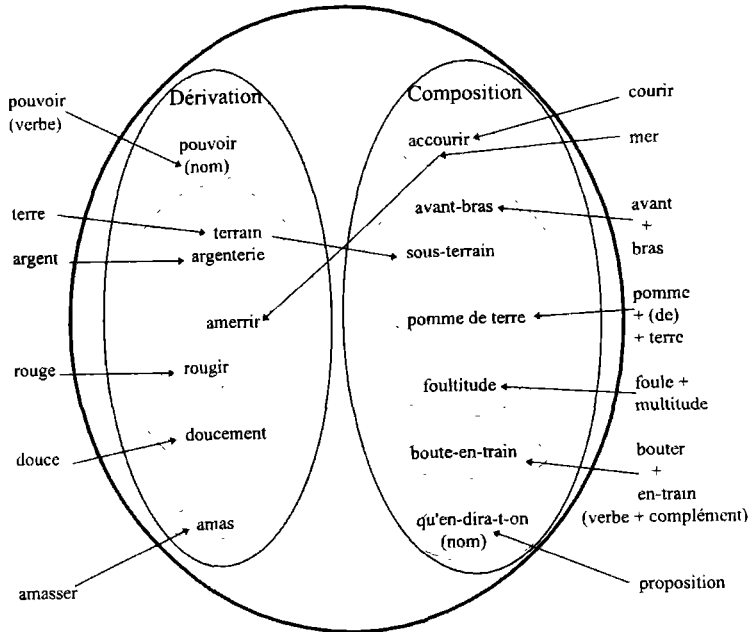
Revoyons rapidement comment fonctionne le système de la morphologie lexicale. Nous parlons en général de dérivation et de composition que l'on peut représenter de la façon suivante (Cardey 1987 : 51-52) :



ou d'une façon un peu plus détaillée faisant apparaître les modes de formation :



Donnons quelques exemples sur sa façon de fonctionner :



En ce qui concerne notre dictionnaire, nous retrouvons certaines formes de dérivations connues mais tout est organisé de façon complètement différente. Prenons un exemple (rappelons que souvent nos terminaisons n'ont rien à voir avec le système dont nous avons l'habitude)

exemple :

nous avons dans notre dictionnaire la règle générale suivante : «les mots terminés par **able** sont des *adjectifs*». Comme exceptions à cette règle générale, nous mettons des *noms* («cartable,...»), des *verbes conjugués* («accable,...»), des *ambiguïtés morphologiques de type nom ou verbe conjugué* («sable, table,...») et des *ambiguïtés de type nom ou adjectif* («comptable,...»).

Un autre exemple, dont la terminaison ne nous est pas familière : (nous ne donnons pas la signification des symboles qui sont bien connus), est donné ci-dessous :

mot se terminant par ague _____	Nom
liste d'exceptions _____	Nom, Verbe conj.
bague; blague; dague; drague	
liste d'exceptions _____	Nom, Adj., Verbe conj.
vague	
liste d'exceptions _____	Verbe conj.
rague; zigzague	

Lorsqu'une terminaison a été adoptée, nous lui avons, puisqu'il est important de reconnaître le plus grand nombre de mots possibles, ajouté les flexions (féminin, pluriel, féminin-pluriel,...).

Ainsi que nous le voyons avec la règle des mots terminés par **ague**, le dictionnaire est constitué d'un ensemble de règles. Chaque règle comprend une condition (la terminaison) et un opérateur (la ou les catégorie(s) grammaticale(s)). Chaque règle générale peut être accompagnée de sous-règles qui peuvent être composées d'une terminaison dépendant de la règle générale (**baire** qui dépend de **aire**) et/ou de listes de mots ayant cette même terminaison («blague, bague» qui dépend de **ague**) associées à un ou des opérateur(s) différent(s) de celui de la règle générale. Cette façon de procéder a donné lieu à une hiérarchisation des sous-règles pour une même règle générale. Les règles générales, elles, pouvant se lire dans n'importe quel ordre.

Soit la règle suivante :

- 21/ Mot se terminant par **al** _____ Adj.
- 21.1/ Mot se terminant par **aal** _____ Nom
- 21.2/ Mot se terminant par **hal** _____ Nom
- 21.2.1/ Consulter la liste _____ Adj.
- catarrhal; nymphal; triomphal; zénithal
- 21.3/ Mot se terminant par **oal** _____ Nom
- 21.4/ Consulter la liste _____ Nom
- amiral; ammonal; arsenal; aval; bacchanal; bal; bocal; cal; canal; caporal; caracal; carnaval; cérémonial; chacal; chenal; cheval; chloral; confessionnal; copal; corporal; dispersal; diurnal; fanal; fécial; festival; foiral; gal; galgal; gavial; géosynclinal; grémial; journal; madrigal; majoral; mémorial; mistral; monial; narval; nopal;

official; orignal; pal; pipéronal; prairial; processionnal; racinal; rational; régal; revival; rorqual; salicional; sandal; serval; sial; signal; sisal; tergal; tincal; tribunal; urinal; val.

21.5/ Consulter la liste _____ Nom, Adj.

animal; anormal; anticlérical; anticlinal; armorial; bancal; brachial; brutal; capital; cardinal; central; clérical; colonial; commensal; cordial; diagonal; dorsal; éditorial; égal; épiscopal; équatorial; final; fromental; général; germinal; international; libéral; littoral; local; méridional; minéral; moral; oral; original; pascal; pectoral; pontifical; présidial; principal; provincial; radical; réal; régional; rival; sidéral; social; spiral; synclinal; vassal; vertical; vespéral.

21.6/ Consulter la liste _____ Nom, Adj., Adv.
mal

Prenons quelques exemples en vue d'expliquer le fonctionnement et surtout l'ordre de lecture des données.

Le mot «cheval» qui se termine par **al** doit pouvoir être traité par notre règle 21. Nous ne devons pas appliquer l'opérateur *Adj.*, il faut d'abord lire les sous-règles. La catégorie grammaticale de «cheval» ne peut pas être définie par les sous-règles 21.1, 21.2, 21.3. La sous-règle 21.4 est une liste à consulter. Le mot s'y trouve, l'opérateur de cette sous-règle, *Nom*, correspond donc à la catégorie grammaticale à laquelle «cheval» appartient.

Pour le mot «matinal», tout comme pour le mot précédent, nous devons parcourir les sous-règles; aucune d'elles ne peut déterminer la nature de ce mot; aussi l'opérateur *Adj.* de la règle générale 21 correspondra à la catégorie à laquelle «matinal» appartient.

Prenons maintenant «trionphal», nous avons une sous-règle qui traite des mots en **hal**, cette dernière comprend elle-même une sous-règle où «trionphal» apparaît; «trionphal» sera donc un *adjectif*.

La formalisation adoptée pour ces règles évite toute confusion possible entre les terminaisons. Cette hiérarchisation de lecture au sein d'une règle est valable tant pour l'humain que pour la machine; ce type de lecture a déjà été utilisé dans certains autres de nos travaux (Cardey et Greenfield 1992).

AVANTAGES D'UN TEL DICTIONNAIRE

Nous avons ici un ensemble de 579 terminaisons accompagnées de sous-terminaisons (sous-règles) et/ou de listes d'exceptions.

Le nombre d'entrées peut paraître important mais essayons de comparer notre dictionnaire avec les autres dictionnaires. Prenons la règle 160 qui reconnaît les formes en **er** (El Harouchy 1997 : 250, tab. 8).

Terminaison er et ses exceptions	Opérateur	Dictionnaire Automatique	Robert
er	Verbe inf.	0	4932
ier	Nom	0	1025
fier	Verbe inf.	0	125
Liste	Nom	3	3
Liste	Adj.	2	2
Liste	Nom, Adj.	1	1
gier	Verbe inf.	0	5
Liste	Nom, Adj.	1	1
Liste	Adj.	42	42
Liste	Nom, Adj.	144	144
Liste	Verbe inf.	127	127
Liste	Nom, Verbe inf.	5	5
Liste	Adv.	3	3
Liste	Nom	184	184
Liste	Adj.	9	9
Liste	Nom, Verbe inf.	41	41
Liste	Nom, Adj.	13	13
Liste	Verbe inf., Adj.	1	1
Liste	Adv.	1	1
Liste	Nom, Adj., Adv.	1	1
Liste	Nom, Adj., Verbe inf.	1	1
Total		579	6666

Cette comparaison s'effectue d'une manière quelque peu spéciale puisqu'une terminaison traite tous les mots qui, en contexte, ont cette terminaison. Nous ne comptabilisons donc pas ces mots qui sont représentés par une unique terminaison et attribuons le chiffre zéro (nous aurions pu mettre le chiffre 1 puisque la terminaison représente une entrée) à cette terminaison. Les seuls chiffres supérieurs à zéro pour le dictionnaire automatique concernent les listes d'exceptions qui sont donc des entrées à part entière. Pour le dictionnaire automatique, un chiffre proche de zéro indique que la terminaison est bien choisie. Il en va autrement pour les autres dictionnaires où plus le chiffre est élevé, plus le nombre d'entrées du dictionnaire est important; l'exhaustivité dépendant de ce nombre d'entrées.

Le dictionnaire Robert sur support électronique recense 6666 entrées pour les mots **er**, le dictionnaire automatique comptabilise un total de 579 «entrées». Ne parlions-nous pas au départ de gain de temps et d'espace mémoire pour le traitement ?

Un autre avantage de ce dictionnaire est qu'il permet de lister les ambiguïtés morphologiques du français.

Voyons l'exemple suivant

6/ Mot se terminant par **ace** _____ Nom

- 6.1/ Consulter la liste _____ Nom, Verbe conj.
 agace; dédicace; espace; glace; grimace; menace; place; préface; surface; trace
- 6.2/ Consulter la liste _____ Nom, Adj.
 biplace; boniface; contumace; dace; monoplace; rapace
- 6.3/ Consulter la liste _____ Adj.
 coriace; efficace; fugace; inefficace; loquace; perspicace; sagace; salace; tenace; vorace
- 6.4/ Consulter la liste _____ Adj., Adv.
 vivace
- 6.5/ Consulter la liste _____ Verbe conj.
 déglace; délance; déplace; désenlace; efface; enlace; entrelace; lace; remplace; replace; retrace; surglace; verglace

Cette règle générale (règle 6) comprend des exceptions. Certaines des sous-règles ont pour opérateur plusieurs catégories grammaticales. Les sous-règles 6.1, 6.2 et 6.4 vont concerner tous les mots en **ace** qui sont des mots polycatégoriels. Ainsi, nous obtenons tous les mots polycatégoriels de notre langue.

Ce dictionnaire permet donc de définir les catégories grammaticales auxquelles une forme ambiguë est susceptible d'appartenir. Ainsi, nous avons établi des règles et de ce fait des constantes. Ce qui fait de ce dictionnaire un outil.

Le dictionnaire permet une reconnaissance des mots non ambigus et des mots polycatégoriels. Il permet de mettre en relief 69 types d'ambiguïtés morphologiques, parmi lesquelles à titre d'exemple, nous trouvons :

- Nom - Participe passé - Verbe Conjugué* (reçus par exemple)
- Nom - Adverbe - Verbe conjugué* (soit)
- Nom - Adverbe - Adjectif - Verbe conjugué* (ferme, double, trouble)
- Nom - Adjectif - Verbe conjugué - Interjection* (fixe)

Pour terminer, nous donnerons l'exemple d'une phrase traitée par ce dictionnaire

la petite joue au ballon

<i>la</i>	sous-règle 2.12 (<i>Art., Nom, Pro. Pers</i>)
<i>petite</i>	sous-règle 281.4 (<i>Nom, Adj.</i>)
<i>joue</i>	sous-règle 398.1 (<i>Nom, Verbe conj.</i>)
<i>au</i>	sous-règle 475.1.5 (<i>Art.</i>)
<i>ballon</i>	règle 368 (<i>Nom</i>)

CONCLUSION

Nous ne pouvons pas dire ici si le hasard fait parfois bien les choses ou s'il y a une logique sous-jacente dans la formation des mots et plus particulièrement dans le cas de nos terminaisons qui ont pu être mises en correspondance avec les catégories grammaticales. Nous pencherions plutôt, dans le cas de notre dictionnaire, vers la première hypothèse sauf pour les cas bien connus des linguistes. Mais une chose assez troublante est que nous avons réalisé, au sein de notre équipe de recherche, le même type de dictionnaire capable

également de déterminer par la terminaison des noms de la langue française si un nom est féminin ou masculin. Il existait quelques règles (les noms terminés par le suffixe **-ier**, **-age**, **-illon**,... sont masculins et ceux terminés par le suffixe **-ade**, **-aie**, **-ée**, sont en principe féminins) mais peu nombreuses et incomplètes; aussi sur le même principe que celui qui a fait l'objet de cet article, un autre dictionnaire a été créé qui définit le genre de tous les noms de la langue française.

RÉFÉRENCES

- CARDEY, Sylviane (1987) : *Traitement algorithmique de la grammaire normative du français pour une utilisation automatique et didactique*. Thèse d'État de Lettres, Besançon, 745 p.
- CARDEY, Sylviane et Peter GREENFIELD (1992) : «A Platform for Teaching Natural Language Morphology», *The Ninth International Conference on Technology and Education*, Paris, vol. 3, pp. 1529-1531.
- DUBOIS, Jean et René LAGANE (1995) : *Grammaire*, Livres de Bord. Larousse, coll. «Bordas», Paris, 514 p.
- DUCROT, Oswald et Jean-Marie SCHAEFFER (1995) : *Nouveau dictionnaire encyclopédique des sciences du langage*, Paris. Seuil, 668 p.
- EL HAROUCHY, Zahra (1992) : *Traitement automatique du genre en français*, DEA Sciences du langage, Besançon, 173 p.
- EL HAROUCHY, Zahra (1997) : *Dictionnaire et Grammaire pour le traitement automatique des ambiguïtés morphologiques des mots simples en français*, Thèse de Doctorat en Sciences du langage, Linguistique et Informatique, Université de Franche-Comté, 486 p.
- GREVISSE, Maurice (1988) : *Le bon usage*, 12^e éd., Paris, Duculot, 1768 p.

UN CAS DE MÉTAPHORE : CRÉATIVITÉ LINGUISTIQUE ET RÔLE EN INNOVATION BIOTECHNOLOGIQUE

Myriam BOUVERET

Université Paul Valéry-Montpellier III, France

1. INTRODUCTION

La néologie en terminologie fait l'objet de règles de formation précises. Or, si l'on observe le fonctionnement sémantique de la dénomination spécialisée, on constate que celle-ci n'est pas qu'une unité de nomenclature, mais qu'elle se comporte comme un signe de langue générale. La créativité lexicale en terminologie est donc loin de se réduire à une néologie selon les règles du système, voire à une néologie de formation gréco-latine. La métaphore est un cas de fonctionnement néologique très riche reposant sur la néologie sémantique et plus seulement sur la néologie de langue.

Nous nous proposons d'analyser la création d'une dénomination spécialisée en biotechnologies, fondée sur une métaphore. Nous montrerons comment une innovation technologique sur laquelle repose la création d'une société est née à partir d'une analogie avec le vocabulaire de la panification. Cette étude repose sur le suivi d'une recherche biotechnologique pendant trois ans, retraçant l'évolution du procédé du laboratoire à l'industrie. Deux hypothèses seront examinées. La première envisage le transfert de vocabulaire non comme un emprunt (néologie dénomminative), mais comme une métaphore (néologie sémantique). On montrera qu'il fonctionne en subduction tant au plan du lexique qu'au plan cognitif. La seconde est que la nature du contexte dans le fonctionnement sémantique de cette métaphore est triple : intra-discursive (mot), discursive et extra-linguistique. Outre les connaissances conceptuelles liées à un domaine, les connaissances sont pratiques mais aussi discursives, révélées et activées dans la mise en discours à travers un ensemble de cooccurrents typiques

2. L'ENJEU DES BIOTECHNOLOGIES EN AROMATISATION

Notre objet est d'analyser la naissance d'une dénomination spécialisée, les motivations et le processus de sa création depuis ses débuts. Le laboratoire de biotechnologie au sein duquel nous avons enquêté¹ est spécialisé dans la recherche d'arômes par voie biotechnologique. Depuis 1985, ses activités de recherches fondamentales et appliquées concernent les substances d'arômes. C'est dans ce cadre que s'inscrit la recherche que nous avons étudiée ici sur la molécule beta-ionone à partir de beta-carotène par voie enzymatique. Ces travaux ont donné naissance à une innovation : un réacteur nommé

1 Anonyme pour des raisons de confidentialité.

pétrin, qui fit l'objet d'un dépôt de brevet d'invention. Ce pétrin doit améliorer la production d'arômes naturels par voie biotechnologique. Deux enjeux majeurs ont été le moteur de cette innovation : le label *naturel* et la réduction des coûts pour une production en quantité industrielle. Les arômes élaborés aujourd'hui peuvent comporter jusqu'à quatre-vingts composants différents. Mais la difficulté pour les arômes de l'industrie agro-alimentaire, destinés à être incorporés dans des denrées alimentaires, est d'obtenir un label d'arômes naturels, les plus intéressants sur le marché étant donné l'engouement du consommateur pour le «naturel», mais aussi les plus coûteux en fabrication. À l'heure actuelle, les législations sur les labels aromatiques sont très strictes et visent à réglementer la qualité des matières premières, leur origine et les procédés de fabrication. Selon la législation, cinq types d'arômes sont autorisés : les *arômes naturels*, les *arômes renforcés*, les *substances identiques aux naturelles*, les *substances artificielles* et les *arômes de transformation*. Mais dans les catégories de produits comme les beurres, crèmes, margarines, crèmes glacées, sorbets et boissons non alcoolisées (saufs boissons aux arômes naturels et artificiels), seuls les arômes naturels sont désormais autorisés.

Dans ce laboratoire, la recherche sur la molécule beta-ionone commence dès 1985, dans le cadre d'une thèse de doctorat. Cette molécule obtenue par voie naturelle est très recherchée. Elle est présente dans les produits de consommation comme le thé, les pruneaux, les jus d'abricot et de mangue, certains jus de fruits chauffés, les algues séchées et le tabac. Cette substance est caractéristique de l'arôme de framboise. Produire la molécule beta-ionone par voie naturelle à des rendements intéressants pour l'industrie constitue toujours un défi. Afin de produire 300 g d'ionone, il faut 1 kilo de carotène. On peut obtenir le beta-carotène en concentrant du jus de carotte, mais il faudrait utiliser 558 kg de jus de carotte ou l'acheter directement à environ 2 000 francs le kilo. Le coût de production serait trop élevé. Une première étude a donc été conduite, situant la réaction en milieu liquide principalement composé de farine, d'huile, de concentré de jus de carotte, d'autres ingrédients et de beaucoup d'eau. Ce travail dura trois ans jusqu'en 1988. Mais un milieu trop liquide ne permettait pas d'arriver aux résultats escomptés, et la recherche s'est poursuivie pendant une année, au cours de laquelle des études comparatives ont été menées entre un milieu liquide, c'est-à-dire un mélange d'ingrédients donnant une consistance liquide, et un nouveau milieu où le liquide était évacué : un milieu qui prenait la consistance d'une pâte.

Cette recherche d'un nouveau milieu réactionnel a conduit à changer de réacteur, il a fallu modifier le système traditionnel, connu pour un milieu liquide. Dès 1989 le nouveau système fonctionne. Il permet de produire plusieurs autres molécules naturelles qui sont alors commercialisées comme l'hexanal naturel à odeur/goût de pomme verte donnant aux aliments un «goût de frais». On n'arrive toujours pas aux résultats satisfaisants pour la molécule beta-ionone. Néanmoins, le système élaboré, un réacteur désigné sous le nom de *pétrin*, et le nouveau milieu réactionnel, un milieu *pâteux*, sont une véritable innovation pour la recherche en aromatisation agro-alimentaire et donneront lieu à un dépôt de brevet en 1993. Dès 1989, le système est pressenti comme étant une innovation technologique sur un marché très porteur, celui des arômes naturels. Et à partir de ce procédé une société de production d'arômes naturels est créée. La société produit avec ce système un ensemble de molécules aromatiques naturelles et poursuit les travaux sur le projet du beta-carotène. En 1995, la recherche sur ce thème est toujours en cours et les dénominations autour du procédé et du milieu évoluent en même temps que la réflexion : le milieu n'est plus complètement *pâteux*, mais d'aspect *mousseux*, et le *pétrin* est désormais dénommé

batteur. Parallèlement, des travaux sont conduits dans le laboratoire de biotechnologie afin d'améliorer les performances de l'un des réactants : l'oxygène.

3. COMPARAISON ENTRE LES CHAMPS LEXICAUX DE L'AROMATISATION ET DE LA PANIFICATION

Les recherches sur la molécule beta-ionone à partir de beta-carotène ont commencé en milieu liquide dans le laboratoire de biotechnologies. Afin d'obtenir la production de beta-ionone, l'ensemble des ingrédients sont mélangés et travaillés à l'intérieur d'un réacteur. Il existe plusieurs sortes de réacteurs, différents selon la nature et la quantité du milieu réactionnel. Une fois le mélange incorporé, la réaction est obtenue par une coupure des chaînes de carbones grâce à une oxydation des lipides, instables, qui réagissent en présence de l'oxygène, c'est la co-oxydation. Les molécules instables qui vont permettre cette coupure sont des radicaux libres. L'oxygène est donc un facteur déterminant, de même que sa circulation. L'autre facteur important est le maintien d'une température constante dans le réacteur. Le milieu est à la base composé de farine, d'huile, d'une lipase : la lipoxygénase, de concentré de jus de carotte : le carotène, et de beaucoup d'eau. Mais d'après le directeur du laboratoire, le milieu était noyé dans trop d'eau, ce qui ne convenait pas. Deux raisons ont donc motivé la concentration du milieu : d'une part la réaction en milieu liquide occasionnait une dégradation à terme de l'un des ingrédients. La plupart des travaux connus sur la lipoxygénase ont en effet été réalisés en milieu très hydraté et l'on a pu démontrer que la réaction en milieu liquide est vingt fois supérieure à celle qui est réalisée en milieu concentré. Mais paradoxalement, ce n'est pas un avantage, car le mélange de l'ensemble des réactants est réalisé avec une mobilité accrue, ce qui entraîne la décomposition de certains éléments. Une réaction en milieu concentré en revanche démarre moins vite mais donne de meilleurs résultats pour obtenir les molécules recherchées. Et d'autre part, la seconde raison est qu'un milieu moins liquide permet d'obtenir une plus grande quantité, ce qui rend le procédé financièrement exploitable à échelle industrielle. Le passage d'un milieu liquide à un milieu concentré s'est donc fait par le recours à un ingrédient particulier : la farine de soja, et par diminution d'un autre : l'eau. Le milieu ainsi produit prenait une consistance toute différente de celui de départ. Il devenait visqueux et prenait l'allure d'une pâte. Et c'est dans ce changement de milieu et le changement de réaction que cela a entraîné que se situe l'originalité du procédé, originalité qui l'est encore à l'heure actuelle où ce type de milieu n'existe pas en aromatisation. On connaît le milieu liquide, le milieu semi-liquide, le milieu solide, mais pas le milieu pâteux. Le cheminement que l'on observe donc est le passage d'un milieu liquide à un milieu plus visqueux. Or ce milieu plus visqueux a l'aspect et la consistance d'une pâte et il se travaille comme une pâte. Cette analogie décrite par les auteurs, dans leur récit oral de l'invention du procédé, se retrouve dans le brevet. Afin de comprendre comment est réalisée le transfert de termes du vocabulaire de la panification à ces travaux, on relèvera pour commencer dans le texte du brevet le vocabulaire utilisé par analogie avec celui de la panification avant d'établir une comparaison avec un article sur la panification.

1- Selon le procédé, on effectue la mise en contact d'une source ou préparation de *lipoxygénases* et d'hydroperoxyde-lyase, d'une source naturelle d'acides gras polyinsaturés et d'une source naturelle de carotène, en milieu concentré à *viscosité* élevée, par *malaxage*, une alimentation en oxygène du mélange réactionnel étant effectuée au cours du *malaxage* dans un réacteur où règne une pression au moins égale à la pression atmosphérique.

2- Le procédé de préparation de substances d'arômes par voie d'oxydation enzymatique est essentiellement original en ce que la teneur en eau est relativement faible, de sorte que la réaction de biosynthèse s'effectue en *milieu à viscosité élevée*, donc concentré en composants bioréactionnels. De ce fait, tout facteur qui augmentera le contact enzyme-substrats accélérera le développement des réactions enzymatiques. On favorisera donc les réactions enzymatiques en assurant un *pétrissage* intensifié des divers composants.

3- La réaction s'effectue dans un *malaxeur*, en *milieu concentré à viscosité élevée*, l'alimentation en oxygène du mélange réactionnel étant effectuée au cours du *malaxage* par un apport d'air ou d'oxygène ou d'un mélange air/oxygène dans un *malaxeur-réacteur* où règne une pression au moins égale à la pression atmosphérique.

4- La réaction est conduite en *milieu concentré, à viscosité élevée*, sous agitation, en présence d'air et/ou d'oxygène, sous pression [...] Si le débit est trop important il en résultera des pertes en arômes élevées par entraînement. On injectera par exemple de 1 à 2 l/m. d'air et/ou d'oxygène pour 2 kg de *mélange pâteux*. La pression d'air, d'air et/ou d'oxygène au sein du *réacteur (mélangeur-malaxeur)* pourra aller jusqu'à 8 bars.

5- En ce qui concerne la *viscosité* du milieu *pâteux*, elle a été caractérisée à l'aide de deux types de mesures.

6- La vitesse d'agitation du *mélange pâteux* sera généralement comprise entre 60 et 200 tours.min.

7- Sur cette figure, on a représenté de façon schématique un réacteur au sein duquel s'effectuent de façon avantageuse les opérations d'homogénéisation et de *malaxage* du *mélange pâteux*, d'oxygénation du *mélange visqueux*, ainsi que la bioréaction qui donne lieu à la formation des arômes par voie d'oxydation enzymatique.

8- Un outil tournant 6, assurant les fonctions de *mélangeage* et de *malaxage*, par mouvement planétaire, est entraîné en rotation par un axe secondaire tournant 7. Cet outil tournant 6, qui constitue le dispositif mécanique d'agitation du *milieu visqueux* 8, en assure simultanément l'homogénéisation et l'oxygénation au sein de l'enceinte étanche 1.

9- En atmosphère libre, qui surmonte le *milieu visqueux*, composé d'air, d'un mélange air/oxygène ou d'oxygène pur, sous pression de l'ordre de 1 à 8 bars, est au contact du *milieu visqueux* 8 brassé par l'outil tournant 6.

10- Les effets conjugués de la pression et du renouvellement de l'atmosphère ont pour conséquence de parfaire les conditions d'oxygénation du *milieu visqueux*.

11- On a obtenu un *milieu pâteux*, qui a été *malaxé* à raison de 90 t.min⁻¹, à 25°C.

12- Les meilleures concentrations en ionones, obtenues après une durée d'essai de 26 h, sont de 95mg.kg⁻¹ de *pâte* pour l'alpha-ionone et de 300 mg.kg⁻¹ de *pâte* pour la beta-ionone.

13- La réaction s'effectue dans un *malaxeur*, en *milieu concentré à viscosité élevée*.

14- Procédé de préparation [...] caractérisé en outre par le fait que le milieu bioréactionnel a une *viscosité* telle que la force d'*adhérence* du *milieu pâteux* peut être comprise entre 5N et 8N.

Les contextes 1 à 4 décrivent un *milieu à viscosité élevée* ou un *milieu concentré à viscosité élevée* et le mélange des composants est nommé *mélange pâteux*. Le terme *pétrissage* est alors employé : «On favorisera donc les réactions enzymatiques en assurant un *pétrissage* intensifié des divers composants». L'analogie avec le vocabulaire de la panification est très forte. Progressivement, on arrive à la dénomination de *milieu pâteux* (5), tout d'abord en cooccurrence avec *viscosité* : *la viscosité du milieu pâteux* et à celle de *mélange pâteux* (4 et 6). Puis, tous les exemples de réaction sont décrits avec le syntagme *milieu pâteux* comme en (11) : «on a obtenu un *milieu pâteux*». Le terme *pâte* est ensuite utilisé comme en (12) pour dénommer le produit obtenu dans le réacteur dans chacun des onze essais de réaction. Ainsi, de la page 8 à 16 du brevet décrivant les onze exemples de réaction, c'est-à-dire dans la plus grande partie du brevet, les seuls termes utilisés sont *milieu pâteux*, *malaxé*, *pâte*, *mélangeur-malaxeur*. Ni *visqueux*, ni *concentré* ne sont employés alors que dans le début du brevet, le syntagme *milieu concentré à viscosité élevé* avait été actualisé plusieurs fois, comme synonyme de *pâteux*, évité. Les onze essais de réactions préparées avec des milieux de proportions différentes dans le *réacteur-malaxeur* sont tous décrits de la manière suivante : «On a obtenu un *milieu pâteux*». Ainsi, dans un moment du récit de brevet qui décrit les résultats d'expérience, c'est le terme *pâteux* qui s'affirme comme seul valable. Les quatre dernières pages du brevet, intitulées «Revendications» et établissant les conclusions, posent comme synonymes les deux dénominations : *milieu concentré à viscosité élevée*, syntagme définitionnel et *milieu pâteux*. À l'écrit, les termes utilisés par analogie avec le vocabulaire de la panification sont donc les suivants : *viscosité*, *visqueux*, *pétrissage*, *pâteux*, *malaxage*, *malaxé*.

Le terme *pétrin* n'est pas actualisé à l'écrit, mais il est utilisé à l'oral dans les entretiens :

«Pour notre procédé on a été cherché des termes dans un autre domaine qui était celui un peu de la boulangerie, panification. [...] Notre réacteur, c'est un *pétrin* maintenant pour faire des réactions enzymatiques» (Entretien avec le directeur, 27 janvier 1994).

Dans le brevet, ce n'est pas *pétrin* qui est employé mais un terme plus générique *malaxeur*, dérivé directement de l'opération de *malaxage* décrite à l'aide du verbe *malaxer*. Dans le brevet, c'est donc le paradigme *pâte*, *pâteux*, *pétrir*, *malaxer*, *malaxeur*, plutôt que *pâte*, *pâteux*, *pétrir*, *pétrin* qui est actualisé. Pour décrire le procédé, de nombreux synonymes sont employés : *mélangeur-malaxeur* ou *malaxeur*, *réacteur* ou *réacteur-malaxeur*. Le terme *mélangeur-malaxeur* est un composé dénommant les fonctions du *pétrin* : il mélange et malaxe. *Réacteur* est le terme hyperonyme, dénomination d'appareil plus générale et *réacteur-malaxeur* décrit la spécificité d'un réacteur, c'est un hyponyme. Le néologisme sémantique *pétrin* est donc évité au profit d'une néologie de langue créée à partir d'une dérivation morphologique : *malaxeur*, sur le modèle de *réacteur*, *mélangeur*.

Quelques extraits d'un article sur l'histoire du pain (Godon 1995) nous permettront d'établir la comparaison entre le champ lexical de départ de la métaphore (A) : la panification et celui du milieu d'accueil (B) : l'aromatisation.

1- On préparait ensuite une *pâte* peu consistante par *malaxage* dans de grandes jarres en terre et on la cuisait dans des moules en terre.

2- Plusieurs opérations se succèdent pendant 30 minutes environ, jusqu'à ce que la *pâte* n'*adhère* plus aux mains de l'ouvrier : délayage du levain, *mélange* et brassage avec la farine, parfois incorporation du sel, découpage de la *pâte*, pliage et battage jusqu'à obtention d'une *pâte* bien liée et aérée.

3- L'eau est indispensable pour préparer la *pâte*. Elle mouille la farine, donne la cohésion à la *pâte* et fournit une mobilité nécessaire aux constituants de la farine pour la réalisation des réactions chimiques.

4- L'eau joue un rôle primordial dans la formation de la *pâte*. Au cours du *pétrissage*, par la multiplication des contacts, les molécules d'eau se lient aux divers groupes électriquement chargés des constituants, puis, par son intermédiaire, des liaisons se forment entre les constituants eux-mêmes. La présence d'eau favorise aussi l'établissement, entre les lipides et les protéines, d'interactions hydrophobes dues aux forces de répulsion des chaînes carbonées vis-à-vis des molécules d'eau.

5- La fabrication du pain dure un peu plus de cinq heures. Pendant ce temps, le boulanger agit sur la *pâte* à certains moments bien précis; entre ces opérations, il laisse reposer ou fermenter la *pâte*, selon le stade de la fabrication. On dit qu'il laisse la *pâte* «travailler».

6- La première opération de la panification est le *pétrissage* : celui-ci permet de mélanger les divers constituants, puis de leur donner, par l'établissement de liaisons chimiques, la structure indispensable pour conférer au pain son aspect. La *pâte* devient *élastique* par formation d'un réseau, le gluten, constitué par des protéines de la farine. L'air indispensable à l'action de la levure est emprisonné par l'action mécanique du *pétrin*.

7- Ce dernier se forme au cours du *pétrissage* par l'établissement de diverses liaisons entre les gliadines et les gluténines, pour donner naissance à la masse *viscoélastique* que l'on obtient quand on passe la *pâte* sous un filet d'eau. Les gliadines constituées d'unités de masse moléculaires de 40 000 donnent, seules, une masse extensible, molle et de faible *élasticité* tandis que les gluténines forment un réseau tenace et peu extensible.

8- L'ensemble des *lipides* ne représentent que deux pour cent environ de la farine [...]. L'ensemble des lipides peut être oxydé par la *lipoxygénase*, l'enzyme qui libère des hydroperoxydes favorables aux propriétés mécaniques des *pâtes*, à partir des acides gras insaturés.

On retrouve ici des termes semblables à ceux du brevet d'aromatisation : *pâte*, *malaxage*, *adhère*, *mélange*, *viscosité*, *pétrissage*. Cette association de termes est équivalente à celle du milieu d'aromatisation. On pourrait conclure que ces termes appartiennent et constituent l'isotopie de la panification. Mais en fait, ce sont seulement ceux qui sont en commun avec l'ensemble identifié pour l'aromatisation. Ce que l'on fait apparaître ici, c'est donc uniquement ce qu'il y a de commun entre le milieu de panification et le milieu d'aromatisation. Il semblerait alors que cet ensemble cohérent forme les cooccurrents typiques du *pétrissage*, plus spécifiquement que ceux de la panification.

4. PRODUCTION DE LA MÉTAPHORE

Le premier niveau de réalisation de cette métaphore est donc le mot. Si tous les autres termes peuvent décrire une activité différente, celui de *pétrissage* est bien spécifique de la panification et est lié à l'appareil : le *pétrin*. La métaphore est réalisée dans le discours du brevet par ce terme clé. *Pétrissage* est en effet le terme pivot, «focus» (cf. Black 1962) ou «foyer» (cf. Le Guern 1973) de la métaphore qui permet le passage du champ lexical A à celui du champ lexical B. L'unité lexicale est donc le premier niveau de réalisation de cette métaphore et le passage du milieu A au milieu B se fait grâce à un partage de caractéristiques communes. Mais il ne s'agit pas d'une simple substitution d'un mot à un autre. On peut aller plus loin et constater qu'il s'agit d'un véritable ensemble de termes en commun. La métaphore ne se situe pas seulement dans le cadre du mot, ni dans celui de la phrase. Elle est réalisée dans le cadre du discours et c'est cette entité qui permet d'identifier le processus métaphorique.

Le deuxième niveau de réalisation de cette métaphore est donc le discours. Celui-ci tout entier est porteur de son développement. Elle déborde largement le cadre du mot et l'analogie est sous-jacente à tout le discours du brevet. Les termes tels qu'ils sont employés en cooccurrence dans le texte : *visqueux*, *viscosité*, *pâteux*, *pâte*, *malaxer*, *malaxage*, *malaxeur* sont issus du vocabulaire de la panification. Mais la *viscosité* dans un autre contexte peut être appliquée à un tout autre domaine de l'agro-alimentaire : au chocolat par exemple. Le terme *viscosité* associé à une *pâte* peut s'appliquer aux fromages : on parle de la *viscoélasticité* des pâtes fromagères. En revanche, la consistance trop liquide du chocolat empêche de le *malaxer*, on le *mélange*. De même que la consistance trop solide du fromage empêche de le *mélanger* ou de le *malaxer*. Cette *pâte* spécifique a donc toutes les caractéristiques d'une pâte à pain. Et c'est la cohérence du discours à travers l'association de cooccurents typiques de la pâte à pain qui crée cette isotopie de la panification.

Le troisième niveau de réalisation de la métaphore est extra-linguistique. Elle est déterminée par un partage de savoirs communs, c'est un savoir expérientiel. Au delà du discours, on retrouve aussi des ingrédients majeurs semblables entre les deux milieux : *farine*, *eau* et *lipides*. La particularité se situe donc dans une association d'ingrédients : *farine*, *huile*, *poudre* (*carotène*) et *eau*, qui ont la consistance d'une *pâte*. Et plus encore, certaines applications sont identiques : la présence d'une quantité d'eau justement dosée pour une meilleure circulation des réactants; la farine assure la cohésion; l'air est indispensable dans la réaction. Ces trois éléments sont ceux qui sont déterminants et novateurs dans le procédé d'aromatisation : une réaction aérobie sur un milieu constitué de farine, de carotène, d'huile et de très peu d'eau. Par ailleurs, nous avons vu plus haut que pour les expériences conduites sur le milieu liquide en aromatisation, on avait pu constater que la rapidité de la réaction en début de phase donnait une concentration moindre de la molécule par rapport à un milieu pâteux où le mélange des réactants était plus lent. Ici, une réaction analogue d'amoindrissement du goût et de l'odeur (*odorante*, *goût*) a été constatée lors d'une vitesse de pétrissage augmentée. Un autre point commun existe : dans les deux milieux pâteux, de panification et d'aromatisation, on constate la présence de l'enzyme *lypoxygénase*. Et le constat établi pour le milieu de panification renforce le parallèle avec la réaction du milieu d'aromatisation :

Si l'origine des substances aromatiques du pain est encore imparfaitement connue, il est cependant certain que les enzymes jouent un rôle essentiel dans la formation de divers

produits précurseurs des arômes au cours du pétrissage et de la fermentation : leur excès modifie en effet beaucoup les qualités organoleptiques du pain (Godon 1995 : 39).

Il semblerait donc qu'à travers l'analogie avec la panification, il ne s'agisse pas que d'une question d'emprunt de vocabulaire : celui-ci est la trace linguistique d'une analogie qui tend à se situer bien au-delà. En l'occurrence, l'analogie est aussi celle de la réaction chimique, des ingrédients. Si l'on a pu constater que la métaphore débordait le cadre du mot pour s'étendre au cadre d'un contexte discursif, on doit admettre ici, que le contexte de la métaphore est aussi d'ordre pratique. La métaphore se réalise donc à trois niveaux : le mot, le discours et la pratique. Ce sont les trois niveaux de contexte que l'on peut identifier dans le fonctionnement sémantique de ce processus néologique.

L'analogie établie ici a un rôle de support de réflexion. On peut donc parler aussi de métaphore cognitive. Ainsi, la métaphore du *pétrin* repose sur un processus d'analogie scientifique et celle-ci a un rôle de support de réflexion dans la recherche. La métaphore est utilisée comme un moteur d'innovation. Et en effet, c'est ce qui apparaît clairement dans un chapitre de l'étude de 1995 dans le cadre d'une recherche du laboratoire de biotechnologies visant à améliorer la circulation d'oxygène dans le procédé.

Cette structure a été établie à partir de la structure biochimique de la *pâte à pain*. Une extrapolation nous a semblé possible dans la mesure où les deux milieux ont une composition très semblable : nous y trouvons lipides, protéines, eau (Document interne, 1995).

Non seulement l'analogie apparaît dans le terme *pâte à pain*, mais c'est la réflexion en termes de panification qui a permis d'établir la structure du milieu réactionnel. L'article cité en bibliographie du document confirme d'ailleurs un mécanisme de recherche par analogie avec le milieu réactionnel de la pâte à pain en panification : «Exemple du rôle de la lipoxygénase en panification [...]» (Nicolas et Potus, 1994, *Science des aliments*, pp. 627-642).

Nous avons donc affaire ici à une métaphore qui se réalise dans le cadre du mot et qui s'étend au discours et plus largement aux pratiques, à l'expérience scientifique. On a pu localiser un foyer de la métaphore : le *pétrissage*, établissant le passage discursif à la panification. Mais, au-delà de ce terme pivot, c'est d'une part la cohérence du discours avec l'association des termes *pâte*, *malaxage*, *viscosité*, *mélanger*, identifiés comme cooccurrents typiques du *pétrissage*, et d'autre part l'analogie des ingrédients et des réactions/applications qui sont le support de cette métaphore. Ainsi, ce procédé d'analogie s'apparente à celui d'une métaphore. Il en a toutes les caractéristiques de partage de deux univers permettant le glissement, mais il en a aussi toutes les caractéristiques de dissociation. Comme pour le processus métaphorique, l'analogie n'est pas totale, mais c'est le partage de certains points communs² entre la panification et l'aromatisation qui permet l'activation de celle-ci.

2 Identifiés comme «typiques» par Black (1962) et Lakoff (1980) ou «inhérents» par Le Guern (1973) et Rastier (1987).

CONCLUSION

Du point de vue linguistique, concernant la diffusion, l'impact de cette néologie est pour l'instant moindre. En revanche, sur le plan cognitif, le processus néologique de la dénomination métaphorique constitue un moteur d'innovation très fort. On ne peut pas se contenter d'analyser cette néologie comme un emprunt de vocabulaire avec simple déplacement de sens. La métaphore a entraîné un transfert de termes et a motivé un transfert de technologie. Dans la métaphore du pétrin, on peut dire que bien qu'il s'agisse du même domaine, l'agro-alimentaire, les deux activités de panification et d'aromatisation sont bien distinctes. De plus le référent est différent dans les deux cas. En aromatisation, le pétrin est fermé, possède une double enveloppe et maintient une température constante, caractéristiques qu'il ne possède pas en panification. Comme l'affirme M. Slodzian, le transfert de connaissances n'est pas qu'un simple transfert de vocabulaire :

«L'analyse des seuls procédés de dénomination ne permet pas d'accéder aux modèles de représentation des concepts. Sans un travail de reconstitution du cadre discursif et sémiotique global, il est difficile d'accéder au chaînage des concepts qui structurent un micro-domaine de savoir. Cette approche de la fabrique linguistique des concepts présuppose un recentrage de l'objectif scientifique, le transfert des connaissances ne pouvant être confondu avec un simple transfert de vocabulaires» (Slodzian 1995 : 243).

Parler de deux acceptions différentes du terme *pâte* ou de l'appareil *pétrin* en milieu A et B ne suffit pas pour expliquer le recours au vocabulaire de la panification. Selon nous, la métaphore dégagée ici n'est ni un emprunt, ni une néologie dénomminative, mais elle est une néologie sémantique à part entière. Cette néologie s'apparente à un processus de polysémie avec mécanisme de subduction (cf. Picoche 1984). Elle est réalisée tant au niveau lexical que discursif et pratique, le contexte d'activation de la métaphore se situe donc à trois niveaux, autant linguistiques qu'extra-linguistiques. La métaphore est bien une néologie sémantique. Même si elle n'a pas une importance quantitative dans le renouvellement de vocabulaires spécialisés, son rôle est fondamental dans la construction et l'avancée des connaissances. Elle déplace les champs de recherche et les champs lexicaux. Dans cette perspective, l'étude de la polysémie peut permettre de dégager des pistes concernant le fonctionnement sémantique des langues spécialisées.

RÉFÉRENCES

- BLACK, M. (1962) : *Models and metaphors. Studies in language and philosophy*, Cornell University Press, 267 p.
- BOULANGER, J.-C. (1984) : «Quelques observations sur l'innovation lexicale spontanée et l'innovation lexicale planifiée», *La banque des mots*, n° 27, Paris, CILF, pp. 3-29.
- BOUVERET, M. (1996) : *Néologie et terminologie : production de sens du terme*. Thèse de doctorat, Université Paul Valéry-Montpellier III.
- BRES, J. (1987) : «Une enquête sociolinguistique en milieu minier : corpus et problèmes méthodologiques», *Lengas*, n° 16, éd. Université Paul Valéry-Montpellier, pp. 74-98.
- CORBIN, D. (1987) : *Morphologie dérivationnelle et structuration du lexique*, 2 vol., Tübingen, Niemeyer, 937 p.

- DARMESTER, A. (1979), (1^o éd. 1887, Delagrave) : *La vie des mots étudiés dans leurs significations*, Champs libres.
- FALL, K., LEARD, J.-M. et P. SIBLOT (dir) (1996) : *Polysémie et construction du sens*, Praxiling, Université Paul-Valéry-Montpellier, 193 p.
- FERNEZ-WALCH, S. (1991) : *L'innovation de produit au quotidien en entreprise industrielle*, Thèse de doctorat en Économie industrielle, École des Mines de Paris.
- GARDIN, B. (1974) : «La néologie, aspects socio-linguistiques», in Guilbert (dir), pp. 67-73.
- GAUDIN, F. (1991) : «Terminologie et travail scientifique : mouvement des signes, mouvement des connaissances», in Gaudin et Assal (dir), pp. 111-132.
- GAUDIN, F. et A. ASSAL (dir) (1991) : «Terminologie et sociolinguistique», *Cahiers de linguistique sociale*, n^o 18, CNRS-Université de Rouen, 213 p.
- GAUDIN, F. et M. BOUVERET (sous presse) : «Partage des noms, partage des notions ? Approche sociolinguistique de difficultés terminologiques en situation interdisciplinaire», Thoiron Ph. (dir), *La dénomination*.
- GODON, B. (1995) : «Le pain», *La science*, mars, Dossier hors série, pp. 38-47.
- GUESPIN, L. (1974) : «Néologie et énonciation : problèmes théoriques et méthodologiques», in Guilbert (dir), pp. 74-82.
- GUILBERT, L. (dir.) (1974) : «La néologie lexicale», *Langages*, n^o 36, Paris, Larousse, 128 p.
- GUILBERT, L. (1975) : *La créativité lexicale*, Paris, Larousse, 285 p.
- KITTAY, E. et A. LEHRER (1981) : «Semantic fields and the structure of metaphor», *Studies in language*, 5-1, pp. 31-63.
- KLEIBER, G. (1990) : *La sémantique du prototype. Catégories et sens lexical*, Paris, PUF, coll. «Linguistique nouvelle», 199 p.
- LAFONT, R. (1978) : *Le travail et la langue*, Flammarion, 301 p.
- LAKOFF, G. et M. JOHNSON (1980) : *Metaphors we live by*, The University of Chicago Press, 241 p.
- LATOUR, B. et S. WOOLGAR (1988) : *La vie de laboratoire, la production des faits scientifiques*, La découverte, 299 p. (traduit de l'anglais, *Laboratory life, the construction of scientific facts*, Sage publications Inc, 1979).
- LECOINTRE, S. et D. LEEMAN (dir) (1994) : «Le lexique : construire l'interprétation», *Langue française*, n^o 103, Paris, Larousse, 125 p.
- LE GUERN, M. (1973) : *Sémantique de la métaphore et de la métonymie*, Larousse Université, 125 p.

- LERAT, P. (1995) : *Les langues spécialisées*, Paris, PUF, coll. «Linguistique nouvelle», 198 p.
- LERAT, P. (1996) : «Terminologie et connaissances sémantiques», Actes des Journées de sémantiques lexicales brestoises, à paraître.
- MOLINO, J. (1979) : «Métaphores, modèles et analogies dans les sciences», *Langages*, n° 54, Paris, Didier Larousse, pp. 83-101.
- PICOCHÉ, J. (1984) : «L'utilisation des notions de signifié de puissance et de subduction en lexicologie», *Cahiers de lexicologie*, vol. 1, n° 44, pp. 41-49.
- PICOCHÉ, J. (1986) : *Structures sémantiques du lexique français*, Nathan, 144 p.
- RASTIER, F. (1987) : *Sémantique interprétative*, Paris, PUF.
- RASTIER, F. (1996) : «Le défigement des expressions figées et leur interprétation», in Fall et alii (dir), pp. 17-24.
- SIBLOT, P. (1993) : «De la prototypicalité lexicale à la stéréotypie discursive : la *casbah* des textes français», in *Lieux communs; topoï, stéréotypes, clichés*, Kimé, Plantin C.(éd), pp. 342-354.
- SIBLOT, P. (1995) : «Comme son nom l'indique...» *Nomination et production de sens*, Thèse d'État, Université Paul Valéry-Montpellier III.
- SLODZIAN, M. (1995) : «Transfert de connaissances nouvelles et aménagement terminologique», *Meta*, Montréal, Presses de l'Université de Montréal, pp. 239-243.

LES NOMENCLATURES ONT-ELLES UNE MÉMOIRE... MÉTAPHORIQUE ? LES SUBSTRATS DE LA NOMINATION

Christine PORTELANCE

Université du Québec à Rimouski, Canada

Ce titre annonce une double interrogation concernant, d'une part, l'existence d'un phénomène sémantique diachronique rattaché à un «ensemble de termes employés dans une science ou une technique», phénomène apparenté à une mémoire et, d'autre part, la nature d'une telle mémoire, c'est-à-dire les propriétés qui la définissent. En d'autres termes, nous cherchons à savoir si les nomenclatures conservent de l'expérience passée une sorte de dépôt pouvant agir sur nos représentations linguistiques à venir et si cette reviviscence est de nature métaphorique. La question de la nomination a été abondamment traitée, surtout par le biais d'études sur la référence¹, cette dernière toutefois ne fait pas l'enjeu de notre démarche, comme nous le verrons plus loin. La nomination est ici traitée comme un jeu de langage², comme une interaction entre «expérience-du-monde» et «expérience-du-langage», et c'est donc en tant que phénomène d'usage plutôt qu'en fonction de sa valeur ontologique qu'elle sera examinée.

L'étude de Louis Guilbert sur la formation du vocabulaire de l'aviation nous sert de champ d'observation pour différentes raisons : premièrement, le domaine est plus technique que scientifique, il annonce ainsi la technoscience; deuxièmement, la période étudiée précède la réalisation d'un vol véritable, on peut donc déceler une convergence entre le développement conceptuel, la formation terminologique et la réalisation technique. Nous nous intéressons principalement aux transferts terminologiques d'une sphère d'expérience à une autre et, à cet égard, l'étude de Louis Guilbert s'inscrit dans la foulée de celle de Wexler³, qui déjà soulignait l'importance de ces transferts terminologiques dans la formation d'un vocabulaire. Or c'est précisément le mouvement même de la migration terminologique qui nous semble porteur d'un nouvel éclairage sur la nomination et qui, par conséquent, constitue l'objet réel de notre réflexion.

1 On pense ici aux travaux de G. Frege, de B. Russell, de W.O. Quine et de P.F. Strawson. Pour un résumé de ces travaux, cf. Hottois (1989).

2 Le point de vue est wittgensteinien même si Wittgenstein dans les *Investigations* conteste la place centrale qu'occupe la nomination dans les études sur le langage. Sa démarche s'attache à identifier et à décrire tous les autres jeux de langage. Nous intégrons la réflexion sur les autres jeux pour revenir à la nomination et tenter ainsi de montrer qu'il y a plus qu'une simple opération de désignation de nature indicielle.

3 Peter J. Wexler, *La Formation du vocabulaire des chemins de fer en France (1778-1842)* publiée en 1955.

VERUM IPSUM FACTUM

«Le vrai est le même que le fait⁴». On ne connaît que ce qu'on a fait. Cette maxime du XVIII^e siècle, attribuée à Giambattista Vico⁵, justifie à elle seule le choix d'examiner un produit de la technologie plutôt qu'une théorie scientifique. En choisissant d'étudier une réalisation concrète, nous pouvons alors sans remords éviter de questionner l'antériorité de la réalité sur l'observation, une perspective qui, de surcroît, a été fortement ébranlée depuis les développements de la physique quantique⁶ dans la première moitié du siècle. Évacuation élégante, pourrait-on dire, de la question de la référence et de la valeur de vérité. Mais un tel choix n'a pas pour seul mérite de se soustraire à la querelle subjectivisme *contra* objectivisme, car il correspond à une posture qui s'inscrit en faux contre une lecture manichéenne du monde. À l'instar de Putnam (1984 : 9) pour qui «l'esprit et le monde construisent conjointement l'esprit et le monde», nous avons la conviction que la réalité consiste en un va-et-vient entre le sujet et l'objet et que cette réflexivité est incontournable⁷.

Qui plus est, la terminologie étudiée par Guilbert s'applique alors à un objet encore virtuel. Cette situation, un obstacle dans le cas d'une étude synchronique simultanée, sert plutôt une démarche comme la nôtre puisque non seulement l'objet n'est pas une chimère, on le sait aujourd'hui, mais la concomitance des constructions langagière et aéronautique nous permet un regard sur ce que le sociologue Bruno Latour (1995) appelle la «science (ou la technoscience) en action». Dès lors, il devient possible de conjuguer l'étude du «fait» à celle du «faire»⁸.

LE TRAITÉ DE FORMATION DU VOCABULAIRE DE L'AVIATION DE LOUIS GUILBERT (1965)⁹

Les documents dépouillés par Louis Guilbert, plus de trois cents, proviennent de sources diverses : brevets, correspondances, revues, journaux ainsi que différents ouvrages dont certains de fiction, notamment ceux de Jules Vernes. Les énoncés d'où les unités terminologiques ont été tirées sont réunis dans un recueil et forment le deuxième tome de l'ouvrage. La période étudiée s'étend sur trois décennies, de 1861 à 1891, Guilbert justifie ce choix :

«Cette délimitation repose sur des motifs d'ordre linguistique et d'ordre historique. 1861, c'est la date d'apparition du mot *hélicoptère*, 1891 suit l'année du brevet d'Ader où est apparu pour la première fois le mot *avion*. Sur le plan du mouvement de la société, la période choisie offre un contexte historique, social et économique caractérisé en premier lieu par un puissant essor des techniques et par leur application à la grande industrie, désormais en plein épanouissement.» (9)

4 De *factum, facere* «faire».

5 Citée et commentée par Glaserfeld dans Watzlawick (1988 : 30).

6 Cf. Heisenberg, W. (1971) *Physique et philosophie*, Albin Michel.

7 Voir à ce propos Portelance (1997).

8 Nous utilisons les termes les plus neutres possibles (*construit* et *construction*, *fabriqué* et *fabrication*) sont par ailleurs trop connotés) dans le sens où, en anglais, les termes *made* et *making* auraient été tout à fait appropriés.

9 Cet ouvrage étant le seul de L. Guilbert auquel nous ferons référence, seuls les numéros de pages figureront entre parenthèses.

Il n'est pas facile de faire la synthèse des analyses de Guilbert — un grand nombre d'observations sur un grand nombre de cas —, on examinera les hypothèses posées au départ, les grandes lignes de ses observations¹⁰ et les conclusions qu'il en tire.

Comme point de départ, Louis Guilbert choisit de s'appuyer sur les propos d'Antoine Meillet et de Ferdinand Brunot. Du premier, il retient que l'influence des causes sociales sur les faits linguistiques se fait particulièrement sentir dans «l'innovation apportée au sens des mots»; du second, qu'il n'existe pas de concordance régulière et constante entre l'évolution parallèle du lexique et des «réalités de la vie et de la pensée» (8). Il pose ensuite la question à laquelle toute sa thèse s'efforcera de répondre :

«Quand la causalité historique suscite une création lexicale, celle-ci vient-elle simplement s'intégrer dans une collection de termes ou provoque-t-elle une redistribution des éléments de signification d'un ensemble lié par un réseau de relation ? [...] En bref, peut-on considérer les notions de diachronie et de structure conciliables ?» (8-9)

C'est bien la question de la structuration du lexique, posée à partir du concept saussurien de valeur, qui sert de cadre à la thèse. Au premier chapitre, il tente d'englober toutes les unités relevées dans un seul champ sémantique autour d'un noyau formé par *navigation aérienne*, «carrefour à partir duquel rayonnent toutes les bifurcations du champ sémantique englobant, [...], les concepts et les signes qui ont trait au déplacement des hommes dans l'air à l'aide d'une machine» (31). Il observe qu'un grand nombre de lexèmes à l'intérieur du champ existent déjà dans d'autres champs, mais que l'opération de transfert faite à partir d'un plan de signification stable ne crée pas un autre plan aussi stable. Il termine par une interrogation, cruciale pour le point qui nous intéresse : «*Navigation* est-il le même signe quand il signifie "navigation maritime" et "navigation aérienne" (71). Le chapitre deux examine le champ morpho-sémantique d'*aviation*, un néologisme de forme et, partant, non pertinent dans la perspective de notre étude. Le chapitre suivant est entièrement consacré aux vocabulaires alimentant l'opération de transfert terminologique. Voici les domaines d'activité par ordre d'apparition dans le texte :

Les sciences connexes de l'aéronautique : physique, mathématiques, astronomie, météorologie, mécanique.

Les sciences naturelles : ornithologie, entomologie, anatomie, ichtyologie.

La marine.

Les transports : maritime, ferroviaire.

Les principales observations concernent les moyens linguistiques de transfert (196-202). Les néologismes d'emploi sont classés en deux catégories : les lexèmes simples et les syntagmes nominaux complexes. Des premiers, Guilbert observe que l'intégration au nouveau champ se fait grâce à un réseau d'oppositions distinctives et il ajoute :

«C'est la constance des rapports contextuels ainsi noués qui les distingue des emplois métaphoriques accidentels du discours. C'est aussi leur appartenance à un ensemble lexical d'origine qui leur confère un statut différent de la métaphore épisodique.» (197)

10 En privilégiant, il va de soi, les données pertinentes à notre étude.

Des arguments sur lesquels nous aurons l'occasion de revenir. Il note de plus qu'une opposition de type *navigation maritime/navigation aérienne* provoque souvent le transfert de lexèmes en série comme *hublot* et *hauban*, qui font ainsi double emploi. Les syntagmes forment des unités de significations complexes autonomes qui apparaissent sous différentes formes. Dans le cas d'un emprunt à la physique ou aux mathématiques, le syntagme se compose d'un premier élément emprunté (une notion théorique, par exemple) auquel s'ajoute une expansion nominale, comme dans *surface de sustentation* ou *angle d'attaque*. Dans d'autres cas, le deuxième élément, de par son contenu sémantique, agit comme marqueur de champ et opère ainsi le transfert; c'est ce deuxième procédé, le plus fréquent, qui est présenté par Guilbert comme caractéristique de la formation d'un vocabulaire. Nous reproduisons l'analyse de l'adjectif de transfert *aérien* en ne rapportant, pour des raisons d'intelligibilité, qu'une partie des exemples.

1. Le premier élément est représenté par un terme d'un autre domaine

wagon aérien
navire aérien
dériveur aérien
mer aérienne
touriste aérien
nageoire aérienne
combat naval aérien
cloche à plongeur aérienne
chevaux de halage aériens

2. Le premier élément désigne un concept de vaste compréhension

équilibre aérien
mouvement aérien
direction aérienne
stabilité aérienne +

3. Le premier élément appartient à la langue générale

exercice aérien
scène aérienne
bond aérien
montagne russe aérienne

Dans ce dernier cas, les syntagmes sont intégrés selon le hasard du discours par la fantaisie de l'auteur, Guilbert les associe à la métaphore.

Le dernier chapitre traite du vol et des différentes parties de l'avion, il ne présente qu'un intérêt secondaire dans le cadre de notre étude. Quant aux conclusions, elles s'organisent autour de six thèmes, nous les résumons à partir de deux points de vue : extralinguistique et linguistique.

Selon Guilbert, le plan extralinguistique représente le passage de la théorie à la pratique. À partir d'une accumulation de données issues des sciences (physique, ornithologie, etc.) et des techniques plus anciennes (transport maritime ou ferroviaire), il y a formulation de nouvelles données théoriques vérifiées par l'expérimentation de modèles élémentaires. La recherche technique, dont la progression s'organise selon les lois de

l'histoire, ne régit pas le processus linguistique puisque l'«évolution des plans du signifiant et du signifié¹¹» n'est pas nécessairement concomitante, mais cette recherche sert plutôt d'impulsion au processus. Le terme *aéroplane*, par exemple, désigne tout d'abord un aérostat de forme plane, puis un appareil de type moderne. L'expérience peut donc évoluer sans susciter de nouvelles dénominations et Guilbert de conclure à une relative autonomie du plan notionnel par rapport au plan linguistique. Par ailleurs, les deux types de relations présentant un caractère systémique observable dans un ensemble lexical donné, estime-t-il, sont la structure du contenu sémantique et la structure morpho-lexicale.

Malgré la nouveauté du domaine d'activité, le nombre de néologismes d'emploi dépasse de beaucoup celui des néologismes de formes, mais néanmoins les mutations linguistiques à l'intérieur des différents lexiques particuliers s'épaulent mutuellement et retentissent sur la langue. Il existe, de plus, une forte tendance à la description, comme le montrent les multiples syntagmes nominaux complexes à valeur analytique. Ces syntagmes sont la plupart du temps formés à partir d'un lexème, emprunté à un autre lexique, auquel s'ajoute une ou plusieurs déterminations à valeur descriptive permettant l'intégration au nouveau vocabulaire. Il y a également utilisation de la composition synthétique à partir d'éléments grecs ou latins, comme si le développement de chaque science ou de chaque technique, note-t-il, contribuait à généraliser certains éléments susceptibles de générer des séries analogiques et permettait au mécanisme de composition synthétique de s'installer ainsi graduellement dans la langue. On retrouve également bon nombre de mots de la langue commune, comme *oiseau*, *aile*, *navire*, *appareil*, introduits dans le nouveau domaine par une sorte de métaphore, mais dont la permanence dans le lexique interdirait, selon lui, d'y voir une «création poétique».

En outre, une remarque de C. Bally sur les similitudes entre l'expression littéraire et l'expression technique, en ce qui a trait à la recherche d'objectivité, d'exactitude et au rejet de la banalité de la langue commune, se vérifie, surtout dans la langue des brevets où les inventeurs semblent vouloir souligner un acte de création par un acte de dénomination, comme si la propriété juridique pouvait être mieux «garantie» par une sorte de «propriété linguistique».

Pour Louis Guilbert, le lexique, relais direct entre l'expérience du monde et la langue, présente bien une organisation linguistique, mais cette organisation ne correspond pas directement à la structuration de l'expérience.

LE TRANSFERT TERMINOLOGIQUE

Notre intention n'est pas de faire ici la critique de la méthode ou des analyses de Louis Guilbert, mais bien de discuter et de relier certaines de ses observations, comme le statut sémantique d'une tête de syntagme (produisant une série syntagmatique). Au chapitre un, Guilbert s'interroge certes sur l'identité de *navigation* dans *navigation aérienne* et *navigation maritime*, mais il oublie d'y revenir lorsqu'il décrit, au chapitre trois, de nombreux cas de transfert sans toutefois en dégager toutes les implications. Il observe, en effet, qu'à un élément emprunté s'ajoute une détermination souvent adjectivale. Cependant,

11 Cette «séparation» par Guilbert des deux plans n'apparaît pas très orthodoxe vu le caractère indissociable du *sā* et *sé* dans le signe, illustré par le célèbre exemple du recto verso d'une feuille. Mais peut-être ne faut-il y voir qu'un glissement sémantique involontaire...

en ajoutant *aérien* à *navigation*, on ne fait pas qu'un simple transfert de lexème, on opère une transformation du noyau sémique du lexème emprunté puisque *aérien* fait bel et bien disparaître le sème «eau» de *navigation*, si bien qu'on voit apparaître l'expression *navire marin* (Guilbert 1965 : 199). La navigation sert de notion de référence, de prototype dirait Lakoff (1987), mais ce faisant le lexème acquiert alors un statut de générique et oblige à la création d'un terme spécifique pour désigner ce qu'il désignait avant le transfert. Si l'expression *navire marin* laisse une impression pléonastique, c'est qu'elle n'est pas passée dans la langue au même titre que la série *navigation maritime, navigation aérienne, navigation spatiale*. Une transformation à répétition du noyau sémique de *navigation* permet de nos jours la navigation sur des réseaux informatiques. Il semble bien que ces faits illustrent, on ne peut mieux, une certaine forme de structuration du lexique en terme de valeur. Ce phénomène, non exclusif au domaine technoscientifique, révèle en outre un trait de la nomination.

Vient ensuite la question de la relative autonomie des plans notionnel et linguistique. L'exemple de Guilbert est celui d'*aéroplane*, il lui sert à montrer la non coïncidence de l'expérience linguistique et de l'expérience technique. Cette induction est effectivement possible. Or, en posant le problème différemment, on peut utiliser le même exemple pour illustrer exactement la situation inverse. Et pour cause, **tous** les signes utilisés dans la période étudiée sont de nature instable : plusieurs disparaîtront, certains passeront dans la langue et d'autres, comme *aéroplane*, connaîtront des fluctuations avant de se fixer. Le signe *aéroplane* a bien survécu à l'objet qui l'avait fait naître — reste à savoir si l'objet désigné était réel ou virtuel — et on peut penser que l'existence de la série *aéro* y est pour quelque chose. Comme le fait remarquer Eco (1992 : 111-112), si les êtres humains «instituent et réorganisent les codes, c'est seulement parce que [les codes] existent déjà». Eco voit dans l'invention d'un signe, un signe flou, imprécis, puisque, dit-il, «[l]es inventions ne s'organisent pas selon un système d'oppositions nettes, mais le long d'un continuum gradué [...]». Ainsi, à sa suite, on peut constater que la notion d'atome a bien évolué depuis Démocrite : description conceptuelle après description conceptuelle, le mot *atome* est en quelque sorte marqué par l'histoire, marqué par des étapes successives de nomination et, en ce sens, il arrive que l'étymologie puisse parfois se superposer à l'histoire des sciences telle une généalogie de la nomination.

MÉTAPHORE, ENCORE ET TOUJOURS LA MÉTAPHORE

D'emblée, Guilbert semble écarter la métaphore du processus de nomination. En effet, lorsqu'il mentionne la métaphore dans les questions de transfert, c'est pour la rejeter en arguant, rappelons-le, que «c'est leur appartenance à un ensemble lexical d'origine qui leur confère un statut différent de la métaphore épisodique.» (197). De même, dans les cas d'utilisation de mots courants, comme on l'a vu plus haut, c'est la permanence dans le lexique qui sert d'empêchement. Associant création poétique et métaphore, rien de surprenant qu'il procède de telle sorte. Ces arguments ne reposent sur rien de véritablement étoffé, mis à part l'argument d'autorité¹². Soulignons tout de même l'existence d'une certaine intuition.

12 Il faut dire que la métaphore était pratiquement évacuée d'office du champ d'investigation de la linguistique structurale.

Il existe toute une kyrielle¹³ d'ouvrages et d'articles portant sur la métaphore. Mode pléthorique ou passion, ce trope, mille fois ramené, mis à toutes les sauces, se dispute avec la métonymie et la synecdoque la place centrale de la rhétorique. Rares pourtant sont les analyses qui offrent plus qu'un développement des idées exposées par Aristote. Certains, comme Nietzsche,¹⁴ adoptent une position extrême en avançant que le langage est par essence métaphorique et nient l'autre pôle, celui de l'écart ou de la figure ornementale. Les études sur la métaphore privilégiées sont ici celles qui s'intéressent à la dimension cognitive de la métaphore.

Notre démarche consiste à voir si les transferts terminologiques présents dans la formation de vocabulaires techniques sont de nature métaphorique en les soumettant à l'épreuve de quelques conceptions de la métaphore.

LA PROPOSITION ARISTOTÉLICHIENNE

«La métaphore consiste à donner à une chose un nom qui lui est étranger en glissant du genre à l'espèce ou de l'espèce au genre, ou de l'espèce à l'espèce, ou bien par analogie.» (*Poétique* 1997 : 49)

À première vue, on trouve dans cette définition quatre types de métaphores, mais on a dit souvent qu'Aristote donnait un sens générique au terme *métaphore*, car les deux premiers types correspondent en fait à la synecdoque¹⁵. Eco (1988) propose de voir le passage de l'espèce à l'espèce non pas comme une métaphore à trois termes, à intersection sémique, mais bien comme un rapport analogique, ce qui laisse une seule véritable catégorie de métaphore. Eco (1988 : 158) fait remarquer que le dernier type, comportant quatre termes (A est à B ce que C est à D¹⁶), correspond à la métaphore cognitive puisque, dans ce cas, Aristote expose ce que la métaphore fait connaître alors que dans les autres cas, il explique comment se produit et se comprend la métaphore. L'acquisition de connaissances se fait par une sorte de conflit conceptuel¹⁷ — la coupe de Dionysos ne ressemble pas en effet au bouclier d'Ares — et, ce faisant, on y apprend quelque chose à la fois sur le «métaphorisé» et sur le «métaphorisant». Étant donné l'existence d'un conflit ou incompatibilité, on ne peut affirmer, contrairement à certaines croyances bien ancrées, que l'analogie se fonde sur la ressemblance, elle fait plutôt «surgir»¹⁸ la ressemblance¹⁹, sinon la métaphore par analogie ne pourrait nous dire que ce que l'on sait déjà.

13 Plus de 4000 titres relevés en 1971, cf. Danesi (1988 : 9).

14 Cf. Kremer-Marietti (1993).

15 Quoique la traduction utilisée par Le Guern (1973 : 32) n'en présente que trois, l'analogie étant présentée comme le procédé de passage...

16 L'exemple d'Aristote : la coupe de Dionysos et le bouclier d'Ares.

17 Pour une analyse tropologique à partir de la notion de «conflit conceptuel», cf. Prandi (1992).

18 Le mot *surgir* remplace l'expression *crée* utilisée lors de la communication pour tenir compte de la discussion lancée par G. Kleiber sur la création à partir de «rien» et la préexistence de la «ressemblance». Ce changement ne modifie pas notre réflexion puisque notre posture épistémologique, bien établie au début du texte, rejette un constructivisme pur et dur. Notre point de vue est plutôt «interactionniste» point de vue que l'on peut rapprocher de celui de Robert (1993).

19 Pour l'analogie appliquée à la création de concepts scientifiques, cf. Schlanger (1991) et Halley (1987).

Cette proposition appliquée à *navigation* donne l'analyse suivante : *navigation* (espèce) passe à *navigation aérienne* (espèce) et, par rétroaction, *navigation* (espèce) passe à *navigation* (genre), crée au passage *navigation maritime* (espèce) et permet la série *navigation spatiale, sur internet, etc.* Le procédé est métaphorique. Le lexème *navigation* subit la transposition; il semble toutefois superficiel de n'y voir que la perte d'un sème /eau/. Nous y reviendrons.

LA PHÉNOMÉNOLOGIE OU PHANÉROSCOPIE DE C.S. PEIRCE

Il n'est pas facile d'extirper la métaphore de la sémiotique fondée par Peirce puisque cette théorie présente un système conceptuel complet dont les éléments une fois isolés perdent de leur sens. Par conséquent, la présentation qui suit possède inévitablement un caractère réducteur.

La sémiotique peircéenne ou phanéroscopie veut rendre compte de l'acquisition de connaissances par un sujet vu comme étant à l'intérieur de la pensée et comme utilisant, pour cette acquisition, des signes qui ne forment pas des unités fixes, mais qui correspondent à un moment de la sémiosis. Le mouvement de la sémiosis va de la priméité (la pensée en terme de pure possibilité) à la secondéité (la pensée événementielle actualisant le premier) vers la tercéité (la connaissance régissant le second, la pensée codifiée) dans un mouvement *ad infinitum*. À ces trois paliers, trois ordres logiques : l'abduction, l'induction, et la déduction. L'abduction « suppose quelque chose de différent de ce qui a été observé et fréquemment quelque chose qu'il serait impossible d'observer directement. » (1978 : 223). Autrement dit, devant un nombre élevé de solutions impossibles à vérifier une par une, un choix s'opérera, intuitivement, par plausibilité. L'abduction créatrice en sciences se produit lorsque la vraisemblance se réalise. Dans cet univers sémiotique, un signe se divise en icône, indices et symbole. Une icône (deuxième) ne peut représenter une idée (troisième) qu'en termes de possibilité (premier), mais un signe peut être iconique, il portera alors le nom d'hypoicône. Pour Peirce, une métaphore est une hypoicône « qui représent[e] le caractère représentatif d'un représentant en représentant un parallélisme dans quelque chose d'autre [...] » (1978 : 149). La particularité de cette définition réside dans l'inclusion d'une capacité de représentation dans la représentation, une telle capacité naît dans un parallélisme avec autre chose, c'est donc une capacité latente d'un signe. Une telle métaphore peut s'appliquer au transfert de nomination, mais non à la dénomination, néanmoins le pouvoir explicatif de cette définition est plutôt faible. Par ailleurs, on reconnaît la filiation du continuum proposé par Eco dans le mouvement infini de la sémiosis et, dans cette logique temporelle, l'opération de transfert pourrait correspondre à un saut fait de l'abduction (premier) à la déduction (troisième), c'est ce que nous retiendrons de l'apport de Peirce.

LAKOFF ET JOHNSON

Dans leur ouvrage conjoint (1980), Lakoff et Johnson accordent à la métaphore une place prépondérante dans la compréhension du monde et de soi et postulent qu'elle nous est aussi essentielle que le sens du toucher. Leur champ d'observation est le langage de la vie quotidienne, ce qui peut sembler loin de nos préoccupations. Les métaphores structurent en

partie la vie de tous les jours et se retrouvent abondamment dans le langage littéral²⁰. Le nombre élevé d'occurrences de ces métaphores quotidiennes leur sert d'argument de base pour critiquer ce qu'ils appellent le mythe de l'objectivisme. Cette critique, qui ramène à l'avant-plan l'hypothèse Sapir/Worf, sera développée par Lakoff (1987). À la suite de cette réflexion en collaboration, les deux auteurs publieront en 1987, chacun de leur côté, des ouvrages pertinents pour notre réflexion.

Chez Lakoff (1987), nous examinerons la notion de prototype et la notion de catégorie radiale. Le prototype, dans sa première version, est une notion opérante dans l'analyse des séries syntagmatiques terminologiques. En effet, comme nous l'avons déjà montré (Portelance 1997), dans une série syntagmatique les différentes expansions de la tête de syntagme constituent des distinctions par rapport à un élément central de la série, élément que l'on peut voir comme le prototype d'une catégorie d'objets et il arrive même qu'avec une innovation technique un élément central soit détrôné au profit d'un autre, ce qui a pour effet de modifier en conséquence les syntagmes appartenant à la série. Dans la catégorisation, le recours à un prototype n'est qu'indirectement lié à la métaphore, aussi, cette notion, en permettant une certaine description du transfert terminologique, n'autorise pas pour autant d'y apposer l'étiquette de métaphore.

La catégorie radiale, selon Lakoff, forme un ensemble dont les membres ne présentent pas un trait commun, mais sont liés par une ressemblance de famille, concept emprunté à Wittgenstein (1986 : 148). Rappelons que ce concept avait été initialement proposé pour décrire un réseau complexe d'analogies, analogie d'ensemble comme de détails, et son exemple de référence, tout sauf anodin, était l'ensemble de tous les jeux (jeux de cartes, jeux de dames et... jeux de langage). Dès lors, le prototype n'est plus un élément central de la catégorie, comme dans la première mouture de la théorie du prototype, il est formé de propriétés définissant la ressemblance. Kleiber (1990) discutant de l'évolution de cette théorie qualifie, non sans une pointe d'ironie, cette deuxième définition du prototype de «prototypique». Lakoff se sert des catégories linguistiques du dyirbal comme illustration : dans cette langue, les noms appartenant à une catégorie sont en quelque sorte chaînés les uns aux autres par leur appartenance à un système de croyances mythologiques plutôt que par des traits sémantiques communs.

En examinant, à partir de ce point de vue, la série qui nous concerne, il semble difficile de déterminer un invariant sémantique entre *navigation maritime, aérienne, spatiale* et *sur internet*, puisque le sème /déplacement/ paraît impossible dans le cas de *navigation sur internet*. On pourrait certes avancer que cette dernière expression est métaphorique et que les autres ne le sont pas. À l'examen des transpositions antérieures de *navigation*, ce raisonnement tient du cercle vicieux; pour en sortir et trouver l'analogie, il faut identifier l'élément qui permet la ressemblance. Nous croyons que ce rapport analogique pourrait être l'exploration, en tant que «passage d'un site à un autre», parce que *navigation par internet* survient après *navigation spatiale* qui fait surgir, avant /exploration/, les sèmes /voyage/ et /découverte/. Ainsi, à chaque ajout, la ressemblance de famille se développe et on peut en effet y voir une sorte de lien radial; par conséquent, en vertu du concept saussurien de valeur, l'analyse sémique doit alors être repensée au fur et à mesure que la famille s'agrandit.

²⁰ Leur approche pose un problème de la définition du littéral, sur cette question cf. Nuessel dans Danesi (1988).

Appliquée au développement d'une théorie scientifique, la ressemblance de famille permet de voir une rupture épistémologique comme un éclatement du réseau analogique de ressemblances et justifie le choix par Kuhn (1970 : 74) de l'expression *paradigme* pour désigner les courants scientifiques.

Avec *The Body in the Mind* (1987), Johnson propose de redonner à l'expérience du corps la place qu'il estime lui revenir dans l'élaboration des projections métaphoriques et des schémas²¹ élaborés mentalement. La station du corps, explique-t-il, permet par exemple de projeter la quantité en terme de verticalité : *plus* est haut; *moins* est bas. Se dégagent alors des catégories, telles devant/derrière, intérieur/extérieur, contenant/contenu, etc., qui deviennent, dans une approche aux accents kantien, les données sensibles de base servant à toute conceptualisation et donnent un point de départ à une théorie de l'imagination. Cette posture séduit dans ce qu'elle contient de promesses eu égard à une compréhension de l'être parlant. De plus, cette représentation «incarnée» a dégagé autour de la représentation de l'espace un champ d'étude fécond actuellement fort prisé en sémantique cognitive. En revanche, ce retour à la phénoménologie, on pense à Merleau-Ponty, se radicalise dans les travaux de Varela *et al.* (1991), lui qui, 10 ans auparavant, avançait que notre expérience n'a pas d'autre fondement que l'existence d'une histoire biologique et sociale commune conduisant à une succession d'interprétations (dans Watzlawick 1988 : 344). Appliquée aux nomenclatures, pareille vision leur accorderait une mémoire des mots absolument vertigineuse..., mais impossible à reconstituer.

EN GUISE DE CONCLUSION

L'invention tient en quelque sorte du mystère, de nombreux témoignages parlent même d'illumination²². Ce qui est fascinant dans le mécanisme d'abduction, c'est comment, en brassant du connu, on arrive à faire surgir du «nouveau». À l'instar d'Halleyn (1987) nous croyons que l'imagination scientifique présente une structure rhétorique : la pensée, en passant de l'abduction à la déduction, est métaphorique; en procédant par bond, elle crée une analogie, une idée qui apparaît au sujet vraisemblable. L'étape de vérification valide ou invalide l'hypothèse. Mais il faut se garder d'y voir une analogie simple; il serait plus réaliste, il nous semble, de concevoir un réseau de relations analogiques. Quant au mécanisme déclencheur de l'idée, nous proposons²³ de l'associer à la notion de «voir comme» utilisée par Wittgenstein pour traiter des figures ambiguës, comme celle du lapin-canard (1986 : 326) dont il est dit :«[...] nous l'interprétons et la voyons comme nous l'interprétons». En transposant «voir comme» à la métaphore, non seulement nous installons la pensée abductrice dans une dimension spatio-temporelle, mais nous affirmons que la métaphore cognitive permet simultanément d'affirmer et d'infirmer : nommer ou créer un objet par métaphore permet en effet de dire ce qu'il est **et** ce qu'il n'est pas. En ce sens, l'analogie créée par la métaphore manifeste également un caractère binaire. Dans les cas de transferts terminologiques, c'est donc la nomination plutôt que la dénomination qui est métaphorique; c'est pourquoi, à notre avis, les terminologies spécialisées forment un réservoir de possibilités de «voir comme» et sont

21 Ce terme (*schemata* dans le texte anglais) est repris à Kant. On peut voir dans cette proposition une démarche conceptuelle inverse de celle qui a produit le concept de «schéma corporel».

22 Cf. l'ouvrage de A. Koestler (1964), *The Act of Creation*, New York, MacMillan cité par Schlanger (1991).

23 Une proposition similaire est discutée par Ricoeur (1975).

elles-mêmes, dans certains cas, issues de l'opération «voir comme». Un des meilleurs exemples, à cet égard, se trouve en physique quantique où les caractères attribués aux quarks sont la couleur, la vérité, l'étrangeté, le charme et la beauté... Le nouveau n'est créé qu'à partir d'une idée, et pour que cette idée se transforme en théorie, elle doit passer par l'étape de la diction. En outre, cette mise en langage, étape incontournable, installe une rétroaction sur l'appareil linguistique servant à la nomination.

Par ailleurs, comme le fait remarquer Schlanger (1991), une œuvre théorique devenue culturelle, comme celle de Copernic ou de Kepler, joue quelquefois un rôle paradigmatique et peut, en tant qu'œuvre, devenir un objet d'étude («poétique» dirait Halley 1987). Entre 1861 et 1891, l'aviation n'est encore qu'une fiction, un vieux rêve en voie de se réaliser certes, mais un rêve que l'on peut faire remonter au mythe d'Icare. Cassirer (1946 : 88), traitant du pouvoir de la métaphore, affirme qu'il y a réciprocity entre la forme du langage et celle du mythe, qui se conditionnent et se déterminent l'une l'autre. Par ailleurs, Peirce, en fondant le pragmatisme, pose qu'il n'y a d'utilité ou de rationalité dans les croyances que lorsque ces dernières permettent l'action (1993 : 333). Par la suite, l'aviation non seulement s'actualisera, mais ses développements subséquents font de la «navigation» plus qu'une technique; la navigation devient structure imaginante car, à travers le prisme de la métaphore spatiale, elle permet aujourd'hui de voyager dans un espace virtuel qui se veut l'icône d'un village global.

RÉFÉRENCES

- ARISTOTE (1997) : *Poétique*, trad. du grec par O. Bellevue et S. Auffret, Mille et une nuits, 93p.
- ARISTOTE (1973) : *Rhétorique*, livre III. Paris, Les Belles Lettres, 173 p.
- CASSIRER, E. (1953) : *Language and Myth*, trad. de l'all. par S.K. Langer (1946 pour la 1^{re} éd.) New York, Dover, 103 p.
- DANESI, M. (dir) (1988) : *Metaphor, Communication & Cognition*, Monography series of the Toronto Semiotic Circle, no 2, University of Toronto, 117 p.
- ECO, U. (1992) : *La production des signes*. Paris, Le Livre de poche, 124 p.
- ECO, U. (1988) : *Sémiotique et philosophie du langage*, [1984], trad. de l'ital. par M. Bouzaher, Paris, PUF, 285 p.
- GUILBERT, L. (1965) : *La formation du vocabulaire de l'aviation*. Paris, Larousse, 709 p.
- HALLEY, F. (1987) : *La structure poétique du monde : Copernic, Kepler*. Paris, Seuil, 275 p.
- HOTTOIS, G. (1989) : *Penser la logique*. Bruxelles, De Boeck Université, 273 p.
- JOHNSON, M. (1987) : *The Body in the Mind : the Bodily Basis of Meaning*. The University of Chicago Press, 233 p.
- KLEIBER, G. (1990) : *La sémantique du prototype*. Paris, PUF, 199 p.
- KREMER-MARIETTI, A. (1992) : *Nietzsche et la rhétorique*. Paris, PUF, 269 p.
- KUHN, T.S. (1970) : *La structure des révolutions scientifiques*, trad. de l'anglais par Laure Meyer, Paris, Champs, Flammarion, 284 p.

- LAKOFF, G. (1987) : *Women, Fire and Dangerous Things. What Categories Reveal about the Mind*, 614 p.
- LAKOFF, G. et M. JOHNSON (1980) : *Metaphors we live by*, The University of Chicago Press, 242 p.
- LATOURE, B. (1995) : *La science en action* [1989], trad. de l'angl. par M. Biezunski, Paris, Folio Gallimard, 663 p.
- LE GUERN, M. (1973) : *Sémantique de la métaphore et de la métonymie*, Paris, Larousse, 126p.
- PEIRCE, C.S. (1993) : *À la recherche d'une méthode*, [inédit] trad. de l'angl. et édité par J. Deledalle et M. Balat sous la dir. de G. Deledalle, Presses universitaire de Perpignan, 375p.
- PEIRCE, C.S. (1978) : *Écrits sur le signe*, rassemblés, traduits et commentés par G. Deledalle, Paris, Seuil, 263 p.
- PORTELANCE, C. (1997) : «De la nomination. Catégorisation et syntagmatique», Toronto, *Sémiotique appliquée/Applied Semiotics*,
<http://www.epas.utoronto.ca:8080/french/as-sa/>
- PRANDI, M. (1992) : *Grammaire philosophique des tropes*, Paris, Minuit, 253 p.
- PUTNAM, H. (1984) : *Raison, vérité et Histoire* [1981], trad. de l'angl. A. Gerschenfeld, Paris, Minuit, 242 p.
- RICOEUR, P. (1975) : *La métaphore vive*, Paris, Seuil, 412 p.
- ROBERT, S. (1993) : *Les mécanismes de la découverte scientifique*, Presses de l'Université d'Ottawa, 262 p.
- STENGERS, I. et J. SCHLANGER (1991) : *Les concepts scientifiques* [1988], Folio Gallimard, 190 p.
- VANDELOISE, C. (1986) : *L'espace en français : sémantique des prépositions spatiales*, Paris, Seuil.
- VARELA, F.J., THOMPSON, E. et E. ROSCH (1991) : *The Embodied Mind*, Cambridge, The MIT Press, 308 p.
- WATZLAWICK, P. (dir) (1988) : *L'invention de la réalité* [1981], trad. de l'all. par A.-L. Hacker, Paris, Seuil, 374 p.
- WITTGENSTEIN, L. (1986) : *Tractatus logico-philosophicus* [1908] suivi de *Investigations philosophiques*, trad. de l'all. par P. Klossowski, Paris, TEL Gallimard, 364 p.

CLASSES D'OBJETS ET BASE DE CONNAISSANCES¹

Man-Ghyu PAK

Université de Kwandong, Corée du Sud

1. PRÉSENTATION DU PROBLÈME

Dans le cas particulier de la compréhension de la langue par une machine, ce qui nous intéresse ici, un des points fondamentaux est bien sûr le traitement des informations sémantiques. Or, la détermination du sens des phrases est le plus souvent difficile à cause de l'ambiguïté des mots de la langue (polysémie). Comme la sémantique nous pousse à aborder la relation entre les mots et la réalité, ce ne sont plus des connaissances linguistiques qui nous permettent de comprendre ces phrases, mais nos connaissances extralinguistiques, c'est-à-dire celles qui touchent le monde. Ces connaissances sur la réalité, sur les entités du monde, permettent une appréciation différente sur le sens des phrases.

Considérons les phrases :

- (1) *J'ai envie de porter quelque chose d'à la mode*
- (2) *Elle porte quelque chose d'affriolant*
- (3) *Elle porte celui qui est à rayures*
- (4) *Il porte celle qui est (fripée + élimée)*
- (5) *Elle ne veut porter que ce qui lui va le mieux*
- (6) *Elle va porter quelque chose d'un peu plus (extravagant + original)*

Il semble très difficile de faire deviner au système que le complément d'objet de *porter* signifie un <vêtement> : en effet d'une part ce complément est pronominalisé, d'autre part le verbe *porter* est polysémique. La méthode dont on se sert généralement pour résoudre ce type de problème consiste à construire une base de connaissances de grande envergure qui comprenne des connaissances extralinguistiques, c'est-à-dire des connaissances générales sur le monde, et à la faire utiliser par le système. Mais cette méthode pose bien des problèmes : par exemple, 1) elle ne donne pas toujours des résultats précis parce qu'elle se fonde sur la théorie des probabilités et 2) elle est très coûteuse.

¹ Cette recherche a été financée par la bourse de recherches de l'Université de Kwandong. Je tiens à remercier Gaston Gross et Denis Le Pesant d'avoir lu le manuscrit et de m'avoir donné de précieux conseils.

Dans cet exposé, nous allons montrer que la notion d'*opérateurs appropriés* permet de résoudre ce genre de problème. Les opérateurs appropriés sont des opérateurs (verbes, adjectifs, noms) qui définissent une classe d'objets. Cette notion a été introduite par G. Gross pour décrire les traits sémantiques des noms des langues naturelles. La classe d'objets <vêtements>, par exemple, est définie, en position de **N0**, par les *opérateurs appropriés* tels que «être à la mode», «aller bien à» et «être à rayures». Notre méthode consiste à recourir au dictionnaire des opérateurs appropriés pour lever les ambiguïtés résultant de la pronominalisation : nous croyons que l'on peut considérer les classes d'objets, non seulement comme un système de connaissances linguistiques, mais aussi comme un système de connaissances sur le monde.

Notre travail a pour but de montrer que les opérateurs appropriés permettent de lever les ambiguïtés du pronom et qu'ils peuvent remplacer, au moins en partie, une base de connaissances parce qu'ils contiennent en fait des connaissances sur le monde.

2. LIMITES DES BASES DE CONNAISSANCES

Dans cette section, nous allons faire une évaluation critique des bases de connaissances.

Notons, d'abord, l'inconvénient consistant à accumuler et à gérer des données de taille énorme qui occupent beaucoup de place et qui sont donc très coûteuses. De plus, en réalité, il est impossible de représenter et d'enregistrer toutes les connaissances sur le monde : il est illusoire de développer une machine dotée des mêmes compétences que le cerveau humain.

Cette méthode ne peut s'appliquer qu'à un domaine de spécialité. C'est pourquoi, aujourd'hui, elle n'est utilisée, généralement, que pour la constitution de <systèmes experts>². Mais les phrases que nous donnons en exemples appartiennent non à un domaine de spécialité, mais à la langue de tous les jours, c'est-à-dire qu'elles sont des expressions de base. Ceci nous conduit à dire que les bases de connaissances ne permettent pas de reconnaître le sens des phrases, même fondamentales.

Généralement, une base de connaissances est représentée par des expressions métalinguistiques : par exemple, la règle de production, les langages de primitives, les réseaux sémantiques, etc. La règle de production est du type <si - alors>. Pour la compréhension de la phrase (3), on formule la règle de production suivante :

Si quelque chose est à rayures, alors c'est un vêtement.

Dans un langage de primitives (Schank 1975), les notions sont décomposées en atomes de sens, les primitives sémantiques, dont la combinatoire engendre les sens à représenter. Les réseaux sémantiques (*semantic network*³) sont des graphes étiquetés dont les nœuds

² Un système expert se compose de la base de connaissances et du moteur d'inférence: la première à son tour est constituée de la base de règles, ensemble des connaissances construites par les experts, et de la base de faits, construite par les usagers pour résoudre des problèmes données.

³ On pourra citer comme travail représentatif *Propositional Semantic Network* de Catelfranchi.

représentent des notions (appelées communément *concepts*) et les arcs des relations. À l'origine, les réseaux sémantiques ont été introduits pour tenter de représenter la mémoire associative humaine. Ces méthodes formalistes ont toutes recours à une forme de calcul pour réaliser des inférences et reconnaître des proximités sémantiques ou des relations de paraphrase. Mais le problème est que ces formalismes sont très compliqués à construire et à gérer parce que la structure des nœuds est trop simple. Par conséquent, la recherche d'une expression complexe prend beaucoup de temps.

Le problème le plus grave posé par les bases de connaissances est que celles-ci ne donnent pas toujours de résultats précis du fait qu'elles se fondent sur la théorie des probabilités. Puisqu'elles s'appuient sur des informations non structurées, elles ne peuvent nous donner que l'ordre d'application et la probabilité statistique. Elles sont donc peu fiables.

3. NOTION DE CLASSE D'OBJETS ET D'OPÉRATEURS APPROPRIÉS

Le traitement automatique traditionnel se faisait sans composante sémantique, à savoir sans analyse sémantique. Mais il est aujourd'hui clair que la traduction ne peut pas fonctionner sans traitement des informations sémantiques. La plupart des systèmes se servent de <traits syntactico-sémantiques> tels que *humain, concret, abstrait*, etc. pour lever les ambiguïtés des mots polysémiques. Les traits sont certes utiles : *concret* et *abstrait* permettent de donner une bonne traduction à deux emplois différents du verbe français *porter* : *porter* par *to bear* en cas de complément abstrait et par *to carry* si l'objet est concret, comme on peut le voir :

- avec un objet abstrait dans :
Luc porte un nom français
Luc bears a French name

- avec un nom concret dans :
Luc porte une valise
Luc carries a case

Cependant ces traits sont insuffisants. Ils n'apportent pas les informations indispensables à la reconnaissance automatique des différents emplois de *porter*, dans la mesure où ces traductions ne valent pas pour tous les abstraits et tous les concrets. Par exemple, un *vêtement* réfère à un concret. Le verbe *porter* qui lui est approprié ne se traduit pourtant pas par *to carry* mais par *to wear* :

Il porte des vêtements de tous les jours
He wears clothes for every day

Si le substantif *vêtement* est codé «concret» comme les termes *valise* ou *sac*, on voit que cette seule spécification ne permet pas de sélectionner la bonne traduction en anglais. Pour prédire la traduction *to wear*, il est nécessaire d'ajouter une information sémantique plus précise telle que «vêtement». G. Gross appelle ces nouvelles informations sémantiques des *classes d'objets*. Cette information doit être notée pour chaque substantif

désignant un vêtement. Le code, noté pour chaque élément de la classe, permet à un analyseur de prévoir la reconnaissance (ou la génération) de l'emploi et la bonne traduction.

Le problème est de déterminer la manière de définir les classes d'objets. Leur originalité réside dans le fait que celles-ci ne sont pas seulement définies de façon sémantique mais par les relations syntaxiques qu'elles entretiennent avec une ou plusieurs classes de verbes et d'adjectifs, appelées opérateurs appropriés⁴. Les opérateurs appropriés verbaux de la classe <vêtements> seraient : *porter, enlever, mettre, changer de, cintrer, enfiler, passer, reprendre, retoucher* (en position objet) ou *aller bien, boulocher, tomber bien* (en position sujet), etc., ce qui veut dire que la classe d'objets <vêtements> est définie par le fait d'être sélectionnée à la fois en position **N1** en (7) ou **N0** en (8) :

- (7) Paul porte **N1**
- Paul enlève **N1**
- Paul met **N1**
- Paul change de **N1**
- Paul cintré **N1**
- Paul enfle **N1**
- Paul passe **N1**
- Paul reprend **N1**
- Paul retouche **N1**

- (8) **N0** va bien à *qqn*
- N0** bouloche
- N0** fait des plis
- N0** tombe bien

De la même façon, on peut définir la classe <vêtements> par les opérateurs adjectivaux, (9) ou (10) :

- (9) Paul est accoutré de **N1**
- Paul est boudiné dans **N1**
- Paul est vêtu de **N1**

- (10) **N0** est à la mode
- N0** est à rayures
- N0** est affriolant
- N0** est ajusté
- N0** est élimé
- N0** est seyant

Voici un extrait de la liste des opérateurs qui définissent la classe d'objets <vêtements>⁵ :

⁴ Pour plus d'informations sur le processus de formation d'une classe d'objets et sur la notion d'opérateurs appropriés, voir Le Pesant (1994).

⁵inc : (trait) inanimé concret, <vêtem> : classe d'objets <vêtements>, E : traduction anglaise.

1. Opérateurs verbaux :

a) En position d'objet :

mettre/N0:hum/N1:inc<vêtem>/E:to put on
porter/N0:hum/N1:inc<vêtem>/E:to wear
enlever/N0:hum/N1:inc<vêtem>/E:to take off
changer/N0:hum/de N1:inc<vêtem>/E:to change Poss <vêtem>
cintre/N0:hum/N1:inc<vêtem>/E:to take in at the waist
enfiler/N0:hum/N1:inc<vêtem>/E:to slip on;to put on;to slip over
passer (usure)/N0:hum/N1:inc<vêtem>/E:
passer/N0:hum/N1:inc<vêtem>/E:to slip in
reprendre/N0:hum/N1:inc<vêtem>/E:to alter;to take in;to let out
retoucher/N0:hum/N1:inc<vêtem>/E:to alter
nettoyer à sec/N0:hum/N1:inc<vêtem>/E:to dry-clean
repasser/N0:hum/N1:inc<vêtem>/E:to iron

b) En position de sujet :

aller bien à /N0:inc<vêtem>/N1:hum/E:to fit hum (taille)
boulocher/N0:inc<vêtem>
tomber bien/N0:inc<vêtem>/E:to fall to the feet
faire des plis/N0:inc<vêtem>/E:to have a crease

2. Opérateurs adjectivaux :

a) En position de sujet :

à la mode/N0:inc<vêtem>/E:fashionable
à plis/N0:inc<vêtem>/E:pleated
à rayures/N0:inc<vêtem>/E:striped
affriolant/N0:inc<vêtem:déshabillé>/E:titillating
ajusté/N0:inc<vêtem>/E:fitted
amidonné/N0:inc<vêtem>/E:starched
élimé/N0:inc<vêtem>/E:worn thin
fripé/N0:inc<vêtem>/E:crumpled
seyant/N0:inc<vêtem>/E:becoming;flattering

b) En position d'objet :

accoutré/N0:hum/de N1:inc<vêtem>/E:got up in
attifé/N0:hum/de N1:inc<vêtem>/E:got up in
botté/N0:hum/de N1:inc<vetem:chaus>/E:
boudiné/N0:hum/dans N1:inc<vêtem>/E:strapped into
sanglé/N0:hum/dans N1:inc<vêtem>/E:strapped up tight in
vêtu/N0:hum/de N1:inc<vêtem>/E:dressed in

Nous présentons aussi un extrait de la la liste de la classe d'objets <vêtements> ainsi définie⁶ :

⁶ G : genre, C : classe d'objets, inc : (trait) inanimé concret, D : domaine R : registre, S : spécifications, E : traduction anglaise, vêt du b : vêtement du bas, vêt du h : vêtement du haut, vêt de c : vêtement de corps, vêt coiff : coiffure, vêt chaus : chaussures.

bas/G:nm/C:vêtem du b/D:habill./R:/S:/E:stocking
béret/G:nm/C:vêtem coiff/D:habill./R:/S:/E:beret
bikini/G:nm/C:vêtem de c/D:habill./R:/S:/E:bikini
blouson de cuir/G:nm/C:vêtem du h/D:habill./R:/S:/E:leather jacket
blue jean/G:nm/C:vêtem du b/D:habill./R:/S:/E:jeans
boots/G:nmp/C:vêtem chauss/D:habill./R:/S:/E:boots
botte/G:nf/C:vêtem chauss/D:habill./R:/S:/E:boot
bottine/G:nf/C:vêtem chauss/D:habill./R:/S:/E:ankle-boot
cale on/G:nm/C:vêtem du b/D:habill./R:/S:/E:underpants;US:shorts
casque/G:nm/C:vêtem coiff/D:mil./R:/S:/E:helmet
casquette/G:nf/C:vêtem coiff/D:habill./R:/S:/E:cap
chandail/G:nm/C:vêtem du h/D:habill./R:/S:/E:pullover,sweater
chapeau/G:nm/C:vêtem coiff/D:habill./R:/S:/E:hat
chaussure/G:nf/C:vêtem du b/D:habill./R:/S:/E:sock
chaussure/G:nf/C:vêtem chauss/D:habill./R:/S:/E:shoe
chemise/G:nf/C:vêtem du h/D:habill./R:/S:/E:shirt
corset/G:nm/C:vêtem de c/D:habill./R:/S:/E:corset
costume/G:nm/C:vêtem (masc)/D:habill./R:/S:/E:complet/E:suit
cravate/G:nf/C:vêtem du h/D:habill./R:/S:/E:tie
culotte/G:nf/C:vêtem du b/D:habill./R:/S:/E:enfant/E:pants
gabardine/G:nf/C:vêtem du h/D:habill./R:/S:/E:gabardine
jean/G:nm/C:vêtem du b/D:habill./R:/S:/E:jeans
jupe/G:nf/C:vêtem du b/D:habill./R:/S:/E:skirt
jupon/G:nm/C:vêtem du b/D:habill./R:/S:/E:(waist) petticoat;US:half slip
manteau/G:nm/C:vêtem du h/D:habill./R:/S:/E:coat;overcoat
pantalon/G:nm/C:vêtem du b/D:habill./R:/S:/E:trousers;US:pants
pardessus/G:nm/C:vêtem du h/D:habill./R:/S:/E:overcoat,topcoat
pyjama/G:nm/C:vêtem/D:habill./R:/S:/E:pyjamas

L'existence de la classe des <vêtements> est justifiée par le fait qu'elle permet la définition des substantifs exigés par beaucoup de verbes et d'adjectifs.

Si on dresse la liste des classes d'objets de cette façon, on pourra faire identifier et reconnaître par le système les différents emplois d'un même verbe polysémique en les décrivant comme suit :

porter <vêtements>: *porter une veste*
porter <marque>: *porter une cicatrice*
porter <arme>: *porter un coutelas*
porter <signe>: *porter un slogan (sur la poitrine)*
porter <enfant>: *porter son enfant en elle (depuis six mois)*
porter <nom>: *porter un drôle de nom*

.....

4. RÉOLUTION SUR LA BASE D'OPÉRATEURS APPROPRIÉS

Nous allons maintenant montrer comment notre méthode résout le problème posé par les phrases (1)-(6). La résolution se fera par les étapes suivantes.

D'abord, le système fait l'analyse morphologique, puis syntaxique du texte donné. Ensuite, il tente de lever l'ambiguïté des mots en se référant au dictionnaire des classes d'objets. Ce travail se fait en deux étapes : D'abord, on étiquette le verbe par ses classes d'objets sur la base du dictionnaire. Ensuite, on reconnaît l'identité de l'acception en combinant le verbe et le substantif (c'est-à-dire le complément d'objet) qui ont en commun la même classe d'objets.

Cependant, lorsque le complément d'objet est un pronom, donc extérieur aux classes d'objets, une reconnaissance de ce type ne marchera pas. Dans ce cas, le système tient compte du type de pronoms. Classiquement, on classe les pronoms en trois catégories :

- les pronoms à usage introductif qui permettent d'introduire un objet ou un ensemble d'objets dans le discours ou dans la conversation (e.g. *quelque chose* dans les phrases (1), (2) et (6));
- les pronoms à usage anaphorique qui renvoient à une entité déjà apparue dans le discours (e.g. *celui* dans la phrase (3), *celle* dans la phrase (4));
- les pronoms à usage déictique qui permettent de désigner directement des entités du monde (*ce* dans la phrase (5); *celui* dans la phrase (3), *celle* dans la phrase selon les cas).

(La distinction entre anaphore et déictique est aujourd'hui très contestée. cf. Kleiber 1990 et 1991)

Pour les pronoms à usage introductif, ils n'ont pas beaucoup de sens en tant que tel. Mais, dans nos exemples, ils obtiennent du sens grâce à leur complément comme d'*à la mode*, d'*affriolant* et d'*un peu plus* (*extravagant + original*). Il faut donc faire reconnaître ce sens par la machine. Dans ce cas, le système recherche toutes les classes d'objets attribuées au verbe en consultant cette fois le dictionnaire des opérateurs appropriés. Si la séquence régie par le pronom en question est reconnue comme un opérateur approprié dans le dictionnaire, le système considère la classe d'objets définie par cet opérateur approprié comme étant le bon référent de ce pronom. S'il s'agit de traduction automatique, le système traduit le verbe par son homologue de la langue cible.

Au cas où le pronom en question est du type anaphorique, il est bien entendu que le système recherche d'abord dans le contexte linguistique des candidats à l'antécédent. Il sélectionne ensuite le candidat approprié, en appliquant la même technique que celle utilisée pour le pronom à usage introductif. (Pour les problèmes généraux de la reconnaissance des anaphores, voir Le Pesant (à paraître)).

Lorsque le pronom en question est du type déictique, le problème est encore plus délicat à traiter : en effet il faut examiner de façon exacte et exhaustive la situation

extralinguistique pour trouver le référent. Mais il existe des cas où les classes d'objets aident à reconnaître le référent. Nous y reviendrons.

Considérons, par exemple, le cas de la reconnaissance de *porter une jupe*. Si l'on fait passer cette séquence dans le dictionnaire des classes d'objets, le substantif *jupe* recevra la marque correspondant à sa classe d'objets de la façon suivante :

(11) jupe : <vêtements>

et le verbe *porter* recevra la marque suivante :

(12) porter1: N1 = <vêtements>
porter2: N1 = <marque>
porter3: N1 = <arme>
porter4: N1 = <signe>
porter5: N1 = <enfant>
porter6: N1 = <nom>

.....

La machine lèvera l'ambiguïté du verbe et reconnaîtra l'identité de l'acception en combinant le substantif et le verbe *porter1*, qui ont en commun le même code de classe d'objets <vêtements>.

Mais lorsque le complément d'objet est un pronom comme (1)-(5), la reconnaissance de ce type ne marche pas. Au cas où le pronom est du type introductif, à la suite de la programmation, le système recherche toutes les classes d'objets attribuées au verbe, ici <vêtements>, <marque>, <arme>, <signe>, <enfant>, <nom>, etc. en consultant cette fois le dictionnaire des opérateurs appropriés. Considérons d'abord le cas des phrases (1) et (2). Rappelons-les pour raisons de commodité:

(1) *J'ai envie de porter quelque chose d'à la mode*
(2) *Elle porte toujours quelque chose d'affriolant*

Si la séquence régie par le pronom en question est trouvée en tant qu'opérateur approprié dans le dictionnaire, ce qui est le cas de (1) et (2), le système nous le montre de la façon suivante :

(1') à la mode/N0:inc<vêtements>
(2') affriolant/N0:inc<vêtement>

Il en tire correctement la conclusion que le pronom *quelque chose* réfère à un vêtement, parce que le sujet (N0) des séquences trouvées, i.e. *à la mode* et *affriolant* est codé de <vêtements>. En traduction automatique, il traduira le verbe par son homologue de la langue cible de la façon suivante :

porter <vêtement>: to wear

Considérons maintenant le cas où le pronom est du type anaphorique, comme dans les phrases (3) et (4). Nous les rappelons ici.

- (3) *Elle porte celui qui est à rayures*
(4) *Il porte celle qui est (fripée + élimée)*

Dans ce cas, le système cherche des candidats pour l'antécédent dans le contexte linguistique. Imaginons qu'il trouve 4 noms masculins singuliers pour candidats à *celui* dans la phrase (3) et 4 noms féminins singuliers pour candidats à *celle* dans les phrases (4). Il procède de la même façon que celle utilisée pour le cas de pronom à usage introductif, c'est-à-dire qu'il recherche maintenant toutes les classes d'objets attribuées au verbe, en consultant le dictionnaire des opérateurs appropriés. Il trouve enfin la séquence régie par *celui* et *celle* en tant qu'opérateur approprié et nous montre le résultat de la façon suivante :

- (3') à rayures/N0:inc<vêtements>
(4') fripé/N0:inc<vêtements>
élimé/N0:inc<vêtements>

Il arrive à reconnaître le sens de chaque pronom en choisissant parmi les 4 noms candidats celui qui est codé par la classe <vêtements>.

Nous avons dit plus haut que, pour les expressions déictiques, le traitement est plus délicat parce que le travail destiné à trouver leurs référents ne peut se réaliser que par la recherche exacte et exhaustive de la situation extralinguistique. Cependant, dans ce cas aussi, les classes d'objets peuvent aider. Considérons, par exemple, la phrase (5) que nous rappelons:

- (5) *Elle ne veut porter que ce qui lui va le mieux*

Ici, on interprétera le pronom *ce* comme une sorte de vêtements grâce à la partie régie par lui. On a en fait

- (5') aller bien à /N0:inc<vêtements>

dans le dictionnaire des opérateurs appropriés de la classe d'objets <vêtements>. Considérons maintenant la phrase:

- Celle-ci fait des plis*

qui peut être interprétée le plus souvent comme 'La jupe est froissée'. Il n'est pas si difficile de reconnaître cette interprétation, parce qu'on trouve

- faire des plis/N0:inc<vêtements>

dans le dictionnaire.

Il peut arriver que si le système ne trouve pas la séquence régie par le pronom, ce qui est le cas de (6) dont *extravagant* (ou *original*) ne figure pas dans le dictionnaire des opérateurs appropriés : en effet, dans cette phrase il y a moins de probabilité que le pronom complément d'objet soit interprété comme *vêtement* que dans les autres phrases. Dans ce cas, le système procède à la méthode de type *préférence* qui permet d'opérer un

arbitrage sur l'ensemble des candidats possibles. Par exemple, ici, on peut avoir pour interprétations du complément d'objet <vêtements>, <marque>, <arme>, <signe>, <enfant>, <nom>, etc., i.e. toutes les classes d'objets avec le verbe *porter*. Si le système a dans l'article de *porter* une liste hiérarchique qui détermine l'ordre d'application de ses classes d'objets, il peut arriver à choisir une interprétation appropriée.

5. CONCLUSION

Nous avons montré que les classes d'objets et les opérateurs appropriés permettent de lever bien des ambiguïtés concernant les pronoms. Pour finir, nous allons souligner les avantages de notre méthode sur celle des bases de connaissances.

- Notons d'abord qu'il n'y a pas à accumuler et à gérer des données de taille énorme, mis à part les données linguistiques. La méthode n'est donc pas très coûteuse.

- Celle-ci ne s'applique pas seulement à un domaine de spécialité, mais à la langue générale, alors que les bases de connaissances ne sont utilisées que pour la constitution de systèmes experts.

- À la différence de bases de connaissances qui s'appuient sur des expressions métalinguistiques, très compliquées à construire, notre méthode se fonde directement sur des expressions de la langue naturelle, ce qui facilite la tâche parce qu'elle n'exige pas un processus compliqué pour l'analyse et la génération du sens.

- Finalement, les bases de connaissances ne donnent pas toujours des résultats précis puisqu'elles se fondent sur la théorie des probabilités. Notre méthode, elle, est basée sur une analyse linguistique stricte.

Nous ne nions pas la nécessité des bases de connaissances, mais seulement nous remarquons que celles-ci doivent se limiter aux domaines de spécialité et que la méthode à base de classes d'objets a pour effet de remplacer une assez grande partie de bases de connaissances.

RÉFÉRENCES

GROSS, Gaston (1992a) : «Forme d'un dictionnaire électronique», *L'environnement traductionnel. La station de travail du traducteur de l'an 2001*, Sillery (Québec), Presses de l'Université du Québec et AUPELF-UREF.

GROSS, Gaston (1992b) : «Classes d'objets et enseignement», *Actes du colloque sur le traitement automatique*, Séoul.

GROSS, Gaston (1994a) : «Classes d'objets et synonymie», *Annales littéraires de l'Université de Besançon*, Série linguistique et sémiotique, vol. 23, Besançon.

GROSS, Gaston (1994b) : «Un outil pour le FLE : les classes d'objets», *Actes du colloque du FLE*, Lille, Presses Universitaires de Lille.

- GROSS, Gaston (1994c) : «Classes d'objets et description des verbes», *Langages*, n° 115, Paris, Larousse.
- GROSS, Gaston (1995) : «Les classes d'objets», *Les concepts opératoires de l'INaLF*, rapport de l'INaLF, pp. 83-88.
- GROSS, Gaston (1996) : *Une sémantique nouvelle pour la traduction automatique : les classes d'objets*, manuscrit.
- KLEIBER, Georges (1990) : «Anaphore-Deixis: deux approches concurrentes», *La Deixis*, Colloque en Sorbonne, 8-9 juin 1990, par M.-A. Morel, L. Danon-Boileau, Paris, PUF, Linguistique Nouvelle.
- KLEIBER, Georges (1991) : «Anaphore-Deixis : où en sommes-nous ?», *L'information grammaticale*, num. 51, octobre 1991, pp. 3-16.
- LE PESANT, Denis (1994) : «Les compléments nominaux du verbe lire, une illustration de la notion de <classe d'objets>», *Langages*, n° 115, Paris, Larousse.
- LE PESANT, Denis (à paraître) : «Anaphores et classes d'objets».
- LEHRBERGER and BOURBEAU (1988) : *Machine Translation: Linguistic Characteristics of MT Systems and General Morphology of Education*, Amsterdam, John Benjamins Publishing Co.
- PAK, Man-ghyu (1993) : *Restructuration de la complétive des verbes de cognition et des verbes de communication en français et en coréen : lexique-grammaire comparé*. Thèse de doctorat, Université Nationale de Séoul.
- PAK, Man-ghyu (1997) : «Traduction automatique et classes d'objets : le problème de porter un vêtement en français et en coréen», *META*, vol. 42, n° 1, mars, Montréal, Presses de l'Université de Montréal, pp. 155-167.
- PUSTEJOVSKY, J. (1993) : «Type Coercion and Lexical Selection», *Semantics and the Lexicon*, Dordrecht, Kluwer Academic Publishers.
- SOWA, J. F. (1993) : «Lexical Structures and Conceptual Structures», *Semantics and the Lexicon*, Dordrecht, Kluwer Academic Publishers.
- SCHANK, R. C. (1975) : *Conceptual Information Processing*, American Elsevier.

ANALYSE DES NOMINALISATIONS DANS UN CORPUS SPÉCIALISÉ : COMPARAISON AVEC LE FONCTIONNEMENT EN CORPUS «GÉNÉRAL»

Anne CONDAMINES

Équipe de Recherche en Syntaxe et Sémantique, UMR 5610 du CNRS, Toulouse, France

INTRODUCTION

En linguistique, un changement important est en cours dans la méthodologie : l'analyse de corpus est en train d'être reconnue comme une approche digne d'intérêt. Les phénomènes langagiers ne sont plus étudiés seulement par introspection mais à travers leur manifestation réelle dans des corpus. Ce mouvement devrait permettre d'établir des liens très étroits entre la lexicologie et la terminologie qui, depuis toujours, met en oeuvre des corpus spécialisés pour construire des listes de termes et même des définitions. Cette convergence de méthodes pourrait avoir des conséquences fructueuses sur l'analyse des corpus spécialisés en terminologie, pour au moins deux raisons. D'une part, l'analyse de corpus spécialisés pourrait bénéficier des développements du point de vue de la définition de méthodes et de la systématisation qui vont avoir lieu en analyse de corpus; deux caractéristiques (méthodologie et systématisme) qui trop souvent font défaut à la terminologie, les données étant recueillies de façon intuitive. D'autre part, pourrait commencer à se mettre en oeuvre une caractérisation des corpus spécialisés, par rapport aux corpus généraux. C'est dans cet objectif d'analyse «différentielle» que se situe l'article. Le fonctionnement des nominalisations dans un corpus spécialisé sur le génie logiciel est décrit par rapport au fonctionnement de ces mêmes nominalisations, dans deux corpus de référence considérés comme généraux. Nous utilisons essentiellement une approche statistique, la plus facile à mettre en oeuvre dans un premier temps, qui nous permet de faire émerger des tendances concernant le fonctionnement des nominalisations.

1. CONTEXTE DE L'ÉTUDE

1.1 Systématisation de l'analyse de corpus en terminologie

Contrairement à la lexicologie, où l'utilisation de corpus est assez récente, surtout en France, la terminologie a toujours fait appel, dans sa pratique, à des textes sur le domaine à l'étude. Sur un domaine spécialisé en effet, il est impossible de faire confiance à son intuition linguistique pour mettre au jour des régularités de fonctionnement permettant d'élaborer un système. Au moment où la linguistique commence à accorder ses lettres de noblesse à l'analyse de corpus, on aurait pu espérer que ce nouveau type d'approche aurait pu bénéficier des acquis de la terminologie en matière d'analyse de corpus. Malheureusement, il n'en est rien; trop souvent en effet, les corpus en terminologie ne

sont utilisés que pour attester d'occurrences, plus rarement pour justifier du choix des termes et encore plus rarement encore pour justifier du choix des relations conceptuelles; les critères qui président à ces choix relèvent encore bien trop souvent de l'intuition.

L'analyse de corpus (entendue au sens d'analyse à l'aide d'outils) pourrait être à l'origine d'une mutation majeure en linguistique ainsi que le pense C. Blanche-Benveniste dans un récent article :

«les concordanciers établis sur des corpus informatisés, en permettant de totaliser les emplois des mots sur des millions de cas, changent totalement la nature de l'analyse distributionnelle. Pour M. Hallyday (1991), les changements sont si considérables qu'on ne peut plus maintenir les grandes oppositions méthodologiques hérités de Saussure» (1997 : 32).

La prise de conscience de l'importance de l'analyse de corpus en linguistique accentue encore la nécessité d'inclure la terminologie dans la linguistique; si les corpus sont considérés comme matériau de base pour l'étude de la «langue générale» alors, il n'y a plus de différence majeure entre lexicologie et terminologie et il faut s'employer à identifier ces convergences méthodologiques. Qui dit corpus dit situation d'énonciation et cette situation peut fort bien correspondre à une activité mettant en oeuvre des connaissances expertes; on est alors exactement dans le type de contexte dans lequel sont réalisées les études en terminologie et en langue spécialisée. Il suffit alors de caractériser le lexique propre au corpus à l'étude, lui-même constitué de façon qu'il soit représentatif d'un discours spécialisé¹, en relation avec une connaissance particulière (voire, une activité particulière); c'est bien ce que soutient Biber (1996 : 171-197), un des chefs de file de l'analyse de corpus.

«In descriptive lexicography, which is concerned with the actual use of words, new meanings are discovered only by examining the use of a word in actual discourse contexts [... In addition, a use perspective is required to investigate the stylistic preferences of individuals, *the differing linguistic preferences of groups of speakers*² and the ways in which "registers" (or "genres") favor some words and structures over others.»

Notre travail se situe dans ce courant qui vise à développer l'analyse de corpus pour la terminologie.

1.2 Mise en place d'une analyse différentielle

Pour mener à bien cette analyse de corpus spécialisés, nous mettons en place une analyse différentielle afin de systématiser l'intuition qui nous semble guider le terminologue, celle qu'un terme est un mot ou un groupe de mots qui fonctionne de manière inattendue, en tout cas d'une manière qui est inconnue du terminologue, non spécialiste du domaine et de la langue spécialisée qui y est à l'oeuvre. Dans ce cas d'ailleurs, plutôt que de «mémoire des mots», qui est évoqué dans le titre du colloque, on

¹ Dans l'article intitulé «Langue de spécialité ou discours spécialisé ?» (1997), nous discutons des différences de contenu entre ces deux appellations.

² C'est nous qui soulignons.

pourrait parler de mémoire du terminologue ! Tout comme le linguiste, le terminologue n'est autre qu'un locuteur qui, par introspection, convoque la mémoire qu'il a des phénomènes langagiers et essaie de retrouver, soit des régularités, soit des irrégularités, pour systématiser les phénomènes langagiers.

On retrouve l'idée de déviance dans différents types de travaux sur la langue. Par exemple, l'étude de la métaphore se fait souvent à partir de cette idée de dysfonctionnement :

«Il est généralement admis que la métaphore repose crucialement sur une 'incorection' ou un 'délit', mais ce trait, une fois dénommé et intégré, donne rarement lieu à une caractérisation satisfaisante» (Kleiber 1994 : 177-206).

Cette même idée de «déviance» a été exploitée dans le projet Eurotra, pour représenter la terminologie par rapport au lexique général, déjà traité. Les sous-langages sont caractérisés de la façon suivante :

«- the restrictive mode : by excluding certain features of GL (general language), SL (Sublanguage) can be described as a restricted form of language;

- the deviant mode : SL can show specific features which are not found in GL and, therefore can be considered a deviant form of GL;

- the preferential mode : this approach of SL phenomena is complementary to the restrictive and deviant modes, and is expressed in terms of preferences» (Mc Naught *et al.* 1991 : 1-7).

On trouve enfin cette idée de déviance dans les études sur la néologie.

«Si un changement intervient dans une règle, il se produit au niveau de la performance, sous la forme d'une déviation, d'une 'faute', et sa transformation en règle nouvelle implique un usage répété, une longue évolution [...] le changement des règles grammaticales échappe à la création consciente. Aucun locuteur, en effet, n'a un comportement linguistique naturel qui le conduit à faire volontairement des fautes [...] Les déviations qui, accumulées, constituent l'usage nouveau échappent à sa volonté, mais créent la règle nouvelle.» (Guilbert 1975 : 29)

L'étude que nous présentons ici nous permet de mettre en oeuvre sur une grande échelle, cette idée de déviance. En effet, elle est faite à partir d'un corpus spécialisé, étudié par rapport à un corpus «général».

1.3 Analyse des nominalisations

Nous avons choisi de faire porter notre attention sur les nominalisations. En effet, d'une part, en tant que noms, elles sont, selon l'approche «traditionnelle» qui privilégie cette forme, considérées comme pouvant être des termes et d'autre part, comme elles sont sémantiquement apparentées à un verbe, leur étude permet une ouverture sur l'étude des actions et événements qui sont trop souvent méconsidérés au profit des seuls objets du domaine. Remarquons, avec plusieurs auteurs, que le nom n'est pas la forme la plus

naturelle d'expression d'une action (on pourrait, à nouveau ici utiliser la notion de «déviance») :

«Il existe aussi des cas de discordance : la nominalisation en est l'illustration. Dans le cas du nom déverbal par exemple, l'action ne trouve pas dans le verbe son expression grammaticale privilégiée mais rencontre une catégorie qui lui est 'étrangère', le nom, et va lui imposer sa combinatoire morphosyntaxique spécifique...» (Rémi-Giraud 1996 : 109).

«La nominalisation — exemplairement le nom d'action — en est l'illustration : par cette métaphore grammaticale, l'action, dégagée de ses attaches référentielles et de l'emprise du temps se trouve mise sur le même plan que les substances autonomes et 'éternelles'» (*ibidem* : 115).

«Comme le nom est réputé représenter un objet, la nominalisation est fort utilisée pour créer un effet d'objectivation : c'est pourquoi elle est massivement attestée dans les textes scientifiques (notamment positivistes) et dans les discours qui les imitent (langue de bois)» (Rastier 1996 : 51).

2. ÉTUDE STATISTIQUE

Les résultats que nous présentons concernent essentiellement des données statistiques. Les études syntaxique et sémantique viendront dans un second temps; leur mise en place sur une grande échelle demande des préparations beaucoup plus lourdes, en particulier, il faudrait au moins disposer d'un corpus automatiquement étiqueté et avec un étiquetage désambiguïsé. Précisons cependant que l'analyse des données statistiques donne une tendance qui permet d'orienter l'analyse sémantique et syntaxique.

2.1 Mise en place de l'étude

2.1.1 Contexte pratique

Le corpus spécialisé qui est utilisé sert de support à un travail mené dans le cadre du GIS (Groupement d'Intérêt Scientifique) Sciences de la Cognition en collaboration avec EDF (Électricité de France), l'IRIT (Institut de Recherche en Informatique de Toulouse) et le LLI (Laboratoire de Linguistique et Informatique). Ce projet vise à affiner un modèle de base de connaissances terminologique (BCT), à définir une méthode assistée par ordinateur de constitution de données terminologiques et à évaluer les possibilités d'utilisation de cette BCT pour constituer un système de consultation de textes. Bien qu'elle ait un lien direct avec ce projet, l'étude qui est présentée ici vise plutôt une réflexion théorique sur le fonctionnement des langues spécialisées. Le corpus, le guide MOUGLIS (Méthodes et Outils de Génie Logiciel pour l'Informatique Scientifique), est constitué par un ensemble de documents méthodologiques sur l'organisation de projets et les techniques de génie logiciel appliquées au développement de logiciels scientifiques à la Direction des Études et Recherches d'EDF. Mougilis est à la fois un guide qui précise comment réaliser les différentes phases de la conception d'un logiciel, et un modèle de rédaction des documents qui accompagnent chacune de ces phases. Il comporte environ 58 000 mots.

2.1.2 Les données étudiées

La première difficulté vient de la sélection des nominalisations déverbiales, qui sont celles qui nous intéressent plus particulièrement. Il est très difficile de définir des critères pour choisir les nominalisations parmi la liste des noms repérés dans le corpus par NOMINO³. Il faudrait combiner à la fois des critères morphologique et des critères sémantiques. Nous avons privilégié, dans un premier temps, une approche assez intuitive qui correspond à la définition donnée par Defranc et Willems : «sous le terme de nom déverbal ou 'nominalisation', nous entendons les formes morphologiquement apparentées aux verbes et ayant en commun avec lui le sème de 'procès'» (1996 : 221).

Parmi les 1151 formes nominales distinctes proposées par Nomino, nous avons retenu 292 nominalisations, soit à peu près le quart. Parmi ces 292 nominalisations, nous n'avons conservé que les 185 formes qui apparaissent au moins trois fois dans le corpus.

2.1.3 Deux corpus de référence

Deux corpus, que nous avons considérés comme «généraux» ont été utilisés comme références : le corpus du TLF (XIX^e et XX^e siècle) et le corpus «*Le Monde*» sur CD-Rom. Bien qu'on puisse toujours mettre en question l'existence d'un corpus «général», c'est-à-dire neutre du point de vue du domaine et du point de vue du style, on peut penser d'une part que la quantité des données permet d'aboutir à une certaine stabilisation des variations et, d'autre part, que le fait d'utiliser deux corpus permet de neutraliser des variations de style. Le corpus du TLF a été mis en oeuvre pour la première caractérisation, le second pour les deux suivantes.

Pour ce qui concerne le TLF, nous l'avons mis en oeuvre via le logiciel HYPERBASE (logiciel construit par É. Brunet (Brunet 1995)) qui propose une fonctionnalité rendant compte de la spécificité du vocabulaire du corpus à l'étude par rapport au vocabulaire du corpus du TLF.

2.1.4 Trois caractéristiques examinées

Trois caractéristiques concernant le fonctionnement statistique des nominalisations sont examinées.

- **Dénombrement d'occurrences**

Il s'agit d'étudier le nombre d'occurrences de chaque nominalisation du corpus spécialisé par rapport au corpus de référence.

Hyperbase fournit une caractérisation statistique, directement utilisable. Trois classes de nominalisations ont été ainsi définies à partir des trois caractérisations proposées par Hyperbase :

³ Outil d'analyse de corpus particulièrement destiné à la terminologie, réalisé par le centre ATO, à la demande du RINT.

- lexèmes en excédent,
- lexèmes en déficit,
- lexèmes absents du modèle.

Les deux premières sont obtenues à partir de «la valeur absolue de l'écart réduit afin de mettre en relief ce qui est le plus significatif»; le troisième concerne les lexèmes absents du fichier REFER qui contient les 10 000 formes les plus fréquentes du TLF (XIX^e et XX^e siècles) c'est-à-dire celles dont la fréquence dépasse 500.

Nous avons considéré une quatrième classe composée des nominalisations qui n'ont aucune spécificité, c'est-à-dire qui fonctionnent comme dans le corpus général.

- Proportion de formes nominales par rapport aux formes verbales

Le corpus du *Monde* nous a permis d'examiner des critères plus complexes.

Le premier critère concerne la proportion de formes verbales par rapport à la forme nominalisée (par exemple, formes verbales de *gérer* par rapport à *gestion*) et ce dans une perspective de comparaison du fonctionnement dans notre corpus par rapport au corpus de référence. Ce premier élément a pour but de vérifier l'hypothèse d'une utilisation de la nominalisation plus élevée dans les corpus spécialisés que dans les corpus généraux.

Pour chaque nominalisation, nous avons calculé le chi² qui met en oeuvre quatre chiffres : nombre de nominalisations dans le corpus, nombre de nominalisations dans *Le Monde*, nombre de formes verbales dans le corpus, nombre de formes verbales dans *Le Monde*. Le chi² est «signé», c'est-à-dire précédé d'un (-) lorsqu'apparaît un déficit de formes nominales par rapport aux formes verbales, dans le corpus.

Une des difficultés que nous avons rencontrées dans le dénombrement est celle des ambiguïtés de formes. En effet, beaucoup de formes de verbes conjugués peuvent être identiques à des formes d'autres catégories (par exemple, *acquis* peut être soit une forme verbale (participe passé ou présent à la deuxième personne ou une forme nominale au pluriel (*les acquis*)). Dans le corpus spécialisé, le nombre d'occurrences est suffisamment faible pour que la vérification systématique des catégories soit possible. En revanche, dans le corpus du *Monde* où il y a parfois des milliers d'occurrences, la vérification est impossible et nous avons fait le choix de ne pas examiner cette caractéristique lorsque des ambiguïtés subsistaient ce qui est encore plus fréquent que pour le corpus spécialisé, dans la mesure où l'interface de recherche du CD-Rom «*Le Monde*» ne tient pas compte des accents (ainsi, la forme *rencontres* correspond soit au nom au pluriel, soit au participe passé du verbe *rencontrer* soit au présent 2^e personne).

Pour le repérage de ces formes verbales, seules les formes correspondant aux temps et aux personnes utilisés dans le corpus spécialisé ont été retenues, soit les 3^{es} personnes du singulier et du pluriel à l'infinitif, au présent et au futur de l'indicatif, le participe passé et le participe présent. En effet, nous avons considéré que le style même du document (un

manuel) plutôt que le domaine imposait ces temps et personnes et qu'il était préférable de comparer les formes verbales sur des bases similaires⁴.

- Proportion de formes nominales au singulier par rapport aux formes nominales au pluriel

Ce critère a pour but de mettre au jour un nombre éventuellement «anormal» de formes nominales au pluriel ou au singulier. Nous voulions tester en particulier l'hypothèse que l'utilisation accrue des nominalisations en corpus spécialisé soit due au fait que l'une des valeurs sémantiques de la nominalisation soit plus fréquente.

On considère généralement qu'une nominalisation peut avoir (au moins) deux valeurs sémantiques : elle peut renvoyer soit au procès, soit au résultat (état ou objet tangible) du procès. Une nominalisation au pluriel renvoie le plus souvent à la deuxième valeur (résultat, objet tangible); ainsi, *les achats*, *les constructions* s'interprètent avec cette deuxième valeur et s'intègrent très mal dans des constructions qui contraignent l'interprétation «processive» :

- *?pendant les achats,
- *pendant les constructions.

Ainsi un nombre significativement élevé de nominalisations au pluriel pouvait correspondre à un nombre important de concepts «objets créés» et contrevenir l'idée que l'utilisation des nominalisations viendrait en lieu et place de l'utilisation des verbes.

2.2 Résultats de l'étude

- Fréquence de nominalisations par rapport au corpus de référence

Les résultats obtenus grâce à Hyperbase sont les suivants :

- 80 nominalisations hors modèle (non comprises dans la liste des formes les plus fréquentes);
- 74 nominalisations en excédent (par rapport au corpus TLF);
- 31 nominalisations conformes au modèle;
- 0 nominalisation en déficit (par rapport au corpus du TLF).

Ces premiers résultats mettent en évidence, d'emblée, une utilisation importante de la nominalisation dans le corpus spécialisé. En effet, 154 nominalisations (80 + 74) apparaissent comme plus fréquentes que dans le corpus de référence, soit plus de 83 %. En revanche, aucune nominalisation n'apparaît en déficit. Cette première évaluation statistique confirme donc, de façon très nette, l'hypothèse d'une utilisation plus fréquente des nominalisations en corpus spécialisé.

Une remarque importante doit cependant être faite. Elle consiste en une mise en garde d'Étienne Brunet lui-même : ces résultats sont à prendre avec des réserves. On peut

⁴ L'étude sur toute les formes verbales, que nous avons aussi faite n'entraîne que très peu de variations.

s'étonner par exemple que ni *listage*, ni *interfaçage* ne soient caractérisés comme plus fréquents dans le corpus spécialisé.

- Nominalisations par rapport aux formes verbales

L'examen des résultats obtenus (avec un autre corpus de référence, je le rappelle) met en évidence une convergence très nette entre les résultats obtenus ici et ceux obtenus avec le critère précédent.

Si l'on considère les résultats correspondant à un chi2 avec un seuil de 0.01 (la probabilité de l'hypothèse nulle est inférieure à 6,63), on constate les éléments suivants :

nominalisations du premier groupe («hors modèle») :

- 27 résultats sur 47 sont supérieurs à 6,63, c'est-à-dire que dans 27 cas sur 47 examinés, le nombre de formes nominales est, de façon significative, supérieur au nombre de formes verbales, par rapport au corpus de référence
- le chi2 moyen est de 80,93

nominalisations du deuxième groupe :

- 20 résultats sur 35 sont supérieurs à 6,63
- le chi2 moyen est de 35,3

nominalisations du troisième groupe :

- 5 résultats sur 18 sont supérieurs à 6,63
- le chi2 moyen est de 4,86 (soit inférieur à celui que nous avons considéré comme significatif : 6,63).

On peut donc dire que, dans l'ensemble, les nominalisations qui n'ont pas une fréquence spécifique dans le corpus ne sont pas utilisées de façon plus élevée que les formes verbales correspondantes. Inversement, pour les deux groupes de nominalisations qui ont une fréquence élevée par rapport au corpus de référence, on constate une utilisation de ces nominalisations bien plus grande que les formes verbales correspondantes. Il y a bien convergence de résultats pour ces deux premières caractérisations : une forme nominale qui est utilisée plus fréquemment dans le corpus spécialisé que dans le corpus général est aussi utilisée plus fréquemment que la forme verbale, toujours par rapport au corpus de référence.

- Nominalisations au pluriel par rapport aux nominalisations au singulier

Dans les trois groupes de verbes, peu de résultats sont supérieurs à 6,63 ou inférieurs à (-)6,63, c'est-à-dire que peu de nominalisations ont une utilisation au pluriel beaucoup plus ou beaucoup moins élevée que dans le corpus de référence :

- 16/60 dans le premier groupe,
 - 13/57 dans le deuxième groupe,
 - 8/25 dans le troisième groupe.
- Soit, au total 37 résultats «anormaux» sur 142 (un peu plus du quart).

Cela signifie que peu de nominalisations ont un comportement déviant, du point de vue du pluriel, par rapport au corpus de référence. L'examen des moyennes de chi2 confirme ces résultats. En effet, dans aucun des trois groupes la moyenne des chi2 est inférieure à (-) 6,63 ou supérieure à 6,63, c'est-à-dire une moyenne qui mettrait en évidence un fonctionnement spécifique :

- groupe 1 :	moyenne des chi2 :	(-),43
- groupe 2 :		0,33
- groupe 3 :		0,67

Notons cependant que, lorsqu'il y a fonctionnement déviant, il va plutôt dans le sens d'une augmentation des formes plurielles (résultat inférieur à (-)6,63) (21 cas sur 37).

Au terme de l'analyse des résultats des trois critères, nous pouvons faire les constatations suivantes. Le corpus spécialisé que nous avons étudié (et sans doute peut-on élargir cette constatation à l'ensemble des corpus spécialisés mais cela reste à vérifier) privilégie les nominalisations par rapport aux formes verbales mais ces nominalisations ne sont pas plus souvent au pluriel que dans les corpus généraux, ce qui tend à montrer que les nominalisations ne sont pas utilisées pour évoquer des objets mais bien en lieu et place des formes verbales.

2.3 Déviance statistique et mise en évidence de termes

La question qui se pose, à l'issue de cette analyse statistique est celle de savoir si on pourrait utiliser, de façon systématique, ce type d'approche (la mise en oeuvre des trois critères) pour repérer les nominalisations qui sont des termes.

Plusieurs raisons nous amènent à penser que cela ne serait pas suffisant pour repérer l'ensemble des termes comportant une nominalisation.

Tout d'abord, il est bien connu que les termes sont, le plus souvent, des syntagmes et non des mots isolés. En conséquence, ce qui pourrait être mis en évidence avec les trois critères, c'est que telle ou telle nominalisation fait partie d'un terme (ou de plusieurs); reste à savoir lequel (ou lesquels) d'autant qu'une nominalisation peut se trouver soit comme tête (ex : *développement logiciel en version définitive*), soit comme expansion (ex : *cycle de développement*).

Ensuite, la mise en oeuvre systématique de ces tests supposent que toutes les données soient disponibles. Or, nous avons vu que pour un certain nombre de cas, les formes verbales peuvent être confondues avec d'autres formes, ce qui empêche la réalisation du test. Pour remédier à cet inconvénient, il faudrait pouvoir disposer de corpus étiquetés sans erreur, ce qui, si on en croit l'article de Habbert et Salem (1995), est loin d'être le cas.

Enfin, nous y reviendrons, l'analyse statistique, seule, ne suffit pas à épuiser les possibilités de repérage de déviance.

Nous ne pensons pas cependant que ces réserves, importantes, ruinent la possibilité d'utiliser les résultats statistiques pour repérer les termes. Il nous faut tout de même mettre en évidence la pertinence des résultats que nous avons obtenus.

Nous avons considéré les nominalisations qui obéissent aux trois critères étudiés : fréquence anormalement élevée dans le corpus, formes nominales significativement plus élevées dans le corpus que dans le corpus de référence, formes nominales au pluriel significativement plus élevées dans le corpus que dans le corpus de référence. Ce dernier critère qui n'est pas caractéristique des corpus spécialisés en général devient pertinent lorsqu'on considère les nominalisations une par une. Rappelons que beaucoup de nominalisations n'ont pas pu subir tous les tests pour des problèmes de catégorisation.

Quinze nominalisations possèdent ces trois caractéristiques de fonctionnement. Treize de ces nominalisations apparaissent dans certains des termes qui ont été retenus à la fin de l'analyse terminologique; ce sont : *configuration*, *évolution*, *fourniture*, *spécification*, *test*, *vérification*, *conception*, *réalisation*, *organisation*, *installation*, *réunion*, *approbation*, *demande*, certains étant particulièrement productifs : *gestion* apparaît dans 43 termes, *test* dans 55, *conception* dans 39, *spécification* dans 18...

Les deux nominalisations qui n'apparaissent dans aucun terme retenu sont *consolidation* et *décision*. *Consolidation* n'a que trois occurrences, ce qui rend les tests statistiques assez peu fiables. Reste *décision* dont rien, dans l'examen des occurrences, laisse penser qu'on pourrait l'intégrer dans un terme.

Ces résultats, somme toute assez pertinents, tendent à prouver que l'approche statistique, à défaut de donner des résultats fermes et définitifs, permet de mettre sur la voie d'éléments terminologiques.

3. AUTRES CRITÈRES DE DÉVIANCE : ÉLÉMENTS SYNTAXIQUES ET SÉMANTIQUES

De nombreux phénomènes, syntaxiques ou sémantiques, pourraient être examinés afin de caractériser le fonctionnement des nominalisations dans un corpus spécialisé. Nous en proposons quelques-uns.

- Une valeur sémantique est préférée

Nous avons déjà rappelé la possibilité, pour une nominalisation, de présenter une polysémie (processus, résultat statif, résultat sous forme d'objet). Il arrive que la langue spécialisée privilégie une de ces valeurs, différente de celle de la langue générale. Par exemple *spécification*, qui, dans la langue générale, renvoie au fait de spécifier, renvoie couramment, dans notre corpus, au contenu d'un document, voire, à ce document même comme le montrent les extraits suivants :

«...non conformité d'un produit ou d'une documentation avec les spécifications du logiciel...»

«...Les spécifications du logiciel définissent en particulier :...»

- Un type de complément est préféré

On trouve souvent évoquée l'ambiguïté de syntagmes comme «le dessin de Paul» qui suppose soit un Paul dessinateur, soit un Paul modèle pour un dessin. Il se peut qu'un

corpus spécialisé, voire la langue spécialisée en général, privilégie certains types de compléments plutôt que d'autres. Ainsi, notre corpus spécialisé fait apparaître très peu de compléments humains pour des nominalisations qui, en théorie du moins, devraient accepter des compléments humains (*la spécification de Paul, le test de Paul...*). Là encore, il faut se garder de se contenter de son intuition sur les potentialités de tel ou tel lexème et privilégier la comparaison avec un corpus de référence afin d'évaluer les différences du corpus spécialisé : analyse très difficile à réaliser automatiquement.

- Le complément peut être éliminé

Nous avons déjà évoqué (Condamines 1995 : 219-238) la possibilité pour un complément d'objet direct, «normalement» obligatoire, de disparaître dans des contextes d'énonciation bien déterminés, ce qui manifestait, pour le verbe, un fonctionnement terminologique. La possibilité d'emploi absolu du verbe serait ainsi l'indice d'une référence implicite car suffisamment partagée par l'ensemble des locuteurs. Ainsi l'énoncé suivant, dans le contexte bancaire : «vous pouvez déposer librement sur votre compte» permet de penser que «déposer» est un terme. Bien qu'elle ne travaille pas sur des langues spécialisées, Michèle Noailly fait à peu de choses près le même genre de constat :

«Par 'emploi absolu' d'un verbe, on entend des emplois où le complément du verbe transitif, direct ou indirect, est absent, sans que cela implique que le verbe en question ait globalement changé de sens [...] cet objet, s'il est nécessairement existant dans l'univers de référence, est linguistiquement considéré comme sans pertinence». (Noailly 1996 : 73-90).

Pour les nominalisations, qui sont beaucoup plus tolérantes aux emplois sans compléments, cette étude est assez difficile à mettre en oeuvre, d'autant qu'elle doit tenir compte des valeurs sémantiques. En tout cas, elle nécessite une comparaison avec le fonctionnement en corpus général avec toujours une difficulté pour mettre en place l'expérimentation.

Même si ces tests ne peuvent être effectués dans un premier temps, leur pertinence peut du moins être évaluée à la main sur quelques exemples. Cette évaluation sera notre prochaine étape, le choix des nominalisations à étudier prendra bien sûr en compte les résultats de l'analyse statistique.

CONCLUSION

L'étude statistique et différentielle des nominalisations d'un corpus spécialisé nous a permis de montrer, grâce à deux tests, la validité d'une hypothèse souvent évoquée pour les corpus spécialisés, celle d'une utilisation plus importante de ce type de formes dans ce type de corpus. Un troisième test nous a permis de montrer que les nominalisations étaient probablement utilisées à la place des formes verbales. Nous avons pu ainsi mettre en oeuvre sur une grande échelle une approche différentielle puisque tous les tests sur le corpus ont été effectués par rapport à des corpus généraux.

En poursuivant dans cette voie de l'analyse différentielle (par rapport à un corpus général, servant de référence), nous espérons arriver à mettre au jour un ensemble de fonctionnements statistiques mais aussi syntaxiques et sémantiques, (malgré la difficulté

pour mettre au point les expérimentations pour ces deux dernières caractérisations) propres aux corpus spécialisés. On peut sans doute espérer que cette caractérisation permettra d'aider au repérage de termes, en rendant un peu plus systématique l'approche intuitive qui sous-tend la construction de terminologies.

RÉFÉRENCES

- BESZTERDA, J. et J. SYPNICKI (1986) : «Les noms d'action en tant que termes», *META*, vol. 32, n° 3, pp. 255-259.
- BIBER, D. (1996) : «Investigating language use through corpus-based analyses of association patterns», *International Journal of Corpus Linguistics*, vol. 1(2), pp. 171-197.
- BLANCHE-BENVENISTE, C. (1996) : «De l'utilité du corpus linguistique», *Revue française de linguistique appliquée*, 1-2, pp. 25-42.
- BRUNET, É. (1996) : «L'hypertexte HYPERBASE», *Lexicomatique et dictionnaires*, Actes des IV^{es} Journées scientifiques du réseau Lexicologie, Terminologie, Traduction de l'AUFELF-UREF, Lyon, 1995, pp. 11-30.
- CONDAMINES, A. (1995) : «Terminology : new needs, new perspectives», *Terminology*, vol. 2.2, pp. 219-238.
- CONDAMINES, A. (1997) : «Langue spécialisée ou discours spécialisé ?», *Mélanges de linguistiques offerts à Rostislav Kocourek*, Halifax, Presses d'ALFA, pp. 171-184.
- DEFRANC, B. et D. WILLEMS (1996) : «De l'abstrait au concret. Une réflexion sur la polysémie des noms déverbaux», *Les noms abstraits, histoire et théorie*, Presses Universitaires du Septentrion, pp. 221-223.
- GUILBERT, L. (1975) : *La créativité lexicale*. Paris, Larousse, collection «langue et langage».
- HABERT, B. et A. SALEM (1995) : «L'utilisation de catégories multiples pour l'analyse quantitative de données textuelles», *TAL (Traitement Automatique des Langues)*, 36-1/2, pp. 249-275.
- KLEIBER, G. (1994) : «Métaphore et déviance : banalisation ou contrainte hiérarchique», *Nominales, essai de sémantique référentielle*, Paris, Armand Colin, pp. 177-206.
- LEBART, L. et A. SALEM (1994) : *Statistique textuelle*, Paris, Dunod.
- NOAILLY, M. (1996) : «Le vide des choses», *Cahiers de Praxématique*, n° 27, pp. 73-90.
- RASTIER, F. (1995) : «Le terme entre ontologie et linguistique», *La Banque des mots, numéro spécial 7, actes des premières rencontres «Terminologie et Intelligence Artificielle»*, pp. 35-64.
- RÉMI-GIRAUD, Sylviane (1996) : «Pour une approche notionnelle de la nominalisation», *Les noms abstraits. Histoire et théorie*, Presses Universitaires du Septentrion, pp. 105-116.
- * Je remercie vivement Max Reinert, statisticien à l'ERSS, qui m'a guidée dans la réalisation de l'analyse statistique.

Annexe : tableau des résultats de l'analyse statistique

- Légende : NC : nominalisations dans le corpus
 NM : Nominalisations dans Le Monde
 VC : verbes dans le corpus
 VM : verbes dans Le Monde
 NsC : Nominalisations au singulier dans le corpus
 NsM : Nominalisations au singulier dans Le Monde
 NpC : Nominalisations au pluriel dans le corpus
 NpM : Nominalisations au pluriel dans Le monde

	NC	NM	VC	VM	chi2 N/V	NsC	NsM	NpC	NpM	chi2 Ns/Np
1er groupe										
achèvement	3	232	15	3413	3,18	3	231	0	1	1,29
adaptation	5	1610	10	3030	(-),1,22	1	1432	4	178	2,35
affinage	4	9	1	268	0,575	4	9	0	0	
ajout	4	87	5	10646	2,0582	4	47	0	40	3,28
appartenance	3	662	2	1404	1,78	3	628	0	34	0,16
approvisionnement	4	364	0	282	3,08	4	307	0	57	0,74
archivage						10	26	0	1	0,38
codage						43	157	0	3	0,81
codification	9	68	0	138	16,83	8	64	1	4	(-),10,35
commande										
compilation	7	181	1	41	0,18	7	128	0	53	2,85
complément						5	380	0	87	1,14
concertation						5	1069	0	32	0,14
configuration	510	255	2	7	7,94	505	226	5	29	10,23
consolidation	3	194	0	1097	16,74	3	189	0	5	7,43
contrôle										
coordination						16	1084	0	32	0,47
couplage						5	31	0	1	
coûts	12	4004	0	2674						
déclenchement	4	274	2	1667	13,46	4	264	0	8	0,12
découpage						19	200	0	26	2,44
définition	69	1512	101	2466	0,45	69	1413	0	99	4,81
démarrage	10	264	0	581	2,74	10	258	0	6	0,23
dépendance	5	578	4	1803	4,75	0	535	5	43	
déroulement	6	539	9	2166	3,75	6	525	0	13	0,14
développement	381	1274	15	5705	10,43	374	8748	7	400	5,74
diffusion	19	4611	2	4533	10,14	19	2812	0	50	0,33
documentation	90	486	19	171	3,7	90	483	0	3	0,55
élaboration	25	805	15	1651	5,33	25	798	0	7	0,21
enregistrement	10	1005	5	4523	5,33	3	659	7	349	(-),5,44
évaluation	6	1338	7	1144	(-),0,31	6	1212	0	126	0,62

A. Condamines

évolution	60	5538	7	1891	8,76	35	4856	25	682	(-146,34)
formalisation	4	16	14	114	1,32	4	13	0	0	
fourniture	37	381	45	2732	81,52	9	286	26	75	(-52,73)
gestion						322	6519	0	27	1,13
identification	37	399	85	1742	10,14	37	389	0	10	0,94
illustration						16	1169	6	332	() 0,33
incorporation	6	77	0	106	14,48	6	76	0	1	(-17,88)
incrementation	4	0	4	0		4	0	0	0	
inspection	10	834	0	108	1,29	10	779	0	61	0,78
intégration						121	2271	0	16	0,65
interaction	7	142	9	18	0,88	1	81	6	61	(-14,92)
livraison						32	576	18	500	2,1
maillage						3	88	0	4	0,13
maintenance						34	318	0	0	1,52
management						9	930	0	1	(-16,67)
modélisation	27	45	1	33	14,19	27	49	0	5	3,22
dommage	3	9	9	6188	1547,74	3	0	0	0	
normalisation	8	648	2	185	3,34	8	640	0	0	
contention	7	304	9	9429	85,37	7	302	0	2	4,63
paramétrage						4	4	0	0	
phasage						3	2	0	0	
pilotage						5	285	0		
planification	15	450	7	231	4,2	15	450	0	4	0,13
position	4	8652	4	175	(-132,84)	4	6668	0	1984	1,16
prestations										
programmation						4	1576	0	37	(-10,42)
prolongement	5	318	9	1344	20,86	5	248	0	70	
prototypage	9	0	0	0		9	0	0	0	
qualification	5	1459	8	4197	1,68	5	1207	0	252	1,04
recommandation	13	730	2	989	11,86	0	198	13	532	(-14,8)
redocumentation	4	0	0	0		4	0	0	0	
référence						140				
regroupement	8	679	20	2413	0,7	8	490	2	189	0,3
recture	12	313	3	241	14,85	9	104	3	9	(-13,62)
réservation						10	229	0	181	(-14,75)
sauvegarde										
schématisation	3	4	0	14		7	3	0	1	0,87
sequencement						4	0	0	0	
spécification	236	41	16	64	127,09	167	6	69	35	(-10,33)
stockage	13	331	5	289	2,49	12	329	1	2	(-1,33)
structuration						12	44	0	0	
support	42	665	21	1510	36,9	36	444	6	221	6,5
test	262	1355	13	742	105,9	68	837	194	518	(-10,26)
traitement						9	2278	21	561	(-10,95)
transfert						22	1272	6	711	2,5
utilisation	55	2081	61	6969	36,21	55	2037	0	44	1,18

Analyse des nominalisations dans un corpus spécialisé : comparaison avec le fonctionnement en corpus «général»

validation						108	97	0	1	1.1
vectorisation	3	0	0			3	0	0	0	
vérification	24	422	22	1438	2177	14	250	10	172	0.73
chi2 moyen					80.33					(-)3.43
2ème groupe										
abstraction						4	397	0	38	0.38
acceptation	23	317	9	7372	34749	23	315	0	2	0.14
accès	13	3962	0	1285						
achat	4	3524	7	2473	-12.27	3	2586	1	938	5.35
acquisition						2	1116	0	443	
analyse										
appel	16	9506	14	10177	0.3	7	8121	9	1385	2.22
application						47	2671	2	453	2.53
appropriation	5	429	5	2134	7.86	4	426	1	3	(-)20.15
approche										
assistance						5	1434	0	5	1.74
association						10	8414	6	3743	4.44
avancement						26	175	0	6	0.88
changement						4	3768	9	1319	(-)12.62
charge										
choix	33	7324	11	8009	12.39					
commentaire	4	2264	3	2148	0.67	3	1371	1	893	0.34
communication						24	515	2	55	0.11
comprehension	7	580	25	6660	8.25	7	578	0	2	2.42
conception	241	2113	13	2526	-234.77	241	1800	0	313	(-)41.17
construction	7	3687	11	4974		7	5185	0	502	0.87
consultation						4	1154	0	538	1.86
contenu										
couverture	5	1201	15	1530	(-)2.79	5	1034	0	157	0.8
décision						14	8938	5	2571	(-)0.17
décision	20	11509	5	12458	10.23	15	8938	5	2571	(-)8.14
decomposition	46	179	8	176	22.88	46	177	0	2	0.51
démarche						38		1		
description	119	1903	140	2261	5.88	114	1758	5	145	1.9
détail						14	1497	1	935	6.37
disposition	21	3684	13	6030	8.17	13	1924	8	1760	0.78
échange										
édition										
enchaînement	7	244	0	888	24.91	5	177	2	67	(-)4.22
étude						8	4888	9	5616	1.87
exécution	11	1630	13	1182	(-)1.43	11	1283	0	347	2.96
exploitation						22	2408	0	154	1.4
extension						6	1283	1	27	(-)5
fabrication						17	1393	0	16	0.19
formation						8	7962	0	1786	1.79

A. Condamines

génération	16	4051	7	498	(-),8,84	16	2731	0	1320	
information						15	6817	18	3984	(-),4,4
	15	2213	3	6360		14	1227	1	986	
	19	1129	12	1917		19	1085	0	44	0,76
justification	4	404	12	3299	3,23	3	318	1		
lecture						16	2596	3	559	4,82
maîtrise						6	2287	0	85	0,22
modification	131	1697	35	3024	3,26,16	81	1108	50	589	(-),0,64
note										
offre										
	91	11436	12	9131	1,56	83	8113	8	3323	
précision	9	1590	41	9678	0,62	6	1191	3	399	(-),0,32
préoccupation	4	1621	3	1241	7,22	0	683	4	938	(-),2,9
préparation	28	1903	15	5256	3,54	28	1858	0	45	0,67
présentation	33	1996	92	16252	30,21	31	1939	2	57	(-),1,18
production						36	8099	1	1082	2,93
proposition	10	6089	68	12497	(-),14,03	9	2819	1	3270	
rappel	7	808	19	7936	9,61	7	681	0	127	1,3
rapport	20	18317	7	2261	(-),6,14	5	15857	15	2460	
réalisation	98	2161	53	10669	239,75	98	1825	0	336	
réception						20	515	0	55	2,12
recherche						7	8944	0	2082	1,62
redaction	58	1683	23	1522	11,56	58	1514	0	169	6,45
remarque										
réponse	19	4790	24	5459	(-),0,11	17	3561	2	1229	2,27
reprise						6				
réunion	34	7189	1	5122	21,58	21	6169	13	1020	15,36
saisie						7		0		
solution	26	5119	0	0						
suppression	3	1806	2	2235	0,47	3	1159	0	647	1,67
surveillance						5	2101	0	25	5,94
synthese	6	965	2	101	(-),2,2	6	899	0	66	0,44
transformation	4	1338	2	3858	5,24	4	1090	0	248	0,9
vision						14	2382	1	300	0,3
chi2 moyen					85,8					0,33
3ème groupe										
abreviation	18	32	0	113	46,2	1	23	17	9	2,03
action						3	9986	43	4601	1,13
arrêt						2	1093	2	327	(-),1,63
assemblage						3	230	1	55	1,41
attribution	4	1420	4	2497	0,65	2	1093	2	327	(-),1,63
calcul	23	563	1	836	30,21	23	849	3	479	5,69
conduite						4	2214	0	342	0,61
connaissance	8	2916	4	8881		8	2174	0	742	2,72
constitution	8	3282	47	8687	(-),4,56	8	3245	0	37	0,1

Analyse des nominalisations dans un corpus spécialisé : comparaison avec le fonctionnement en corpus «général»

contrainte						1	839	28	1285	
convention	6	3214	7	41	(-)22,21	3	2435	3	779	(-)2,16
demande						0		8		
dépense						0		4		
direction						4	11142	0	679	0,24
écriture						5	2111	0	198	0,46
engagement	4	4632	2	9751	3,26	3	2852	1	1780	0,11
essai	5	6298	0	2785	2,21	0	1800	5	4498	(-)2
existence	6	3691	7	6900	0,73	6	3619	0	72	0,11
expression	5	3349	6	5989	0,43	5	3063	0	286	0,46
fonctionnement	6	2518	0	1973	4,69	6	2500	0	18	4,31
instruction	4	3799	0	585	0,61	0	3377	4	422	(-)3,74
intérêt	4	11806	5	5117	(-)2,73	4	8292	0	3514	1,69
interfaçage	2	0	0			2	0	0	0	
intervention	4	5087	3	3128	(-)16,77	3	3965	1	1122	(-)2,01
listage						2	1	2	0	
mesure										
opération						8	6248	3	3666	0,44
personnalisation	7	45	7	256	(-)1,92	7	45	0	0	
question	11	17562	0	167	0,1	5	12466	6	5096	(-)3,47
signification	4	561	0	1629	(-)1,59	4	512	0	49	0,38
travail	17	22454	4	7867	0,51	17	22453	0	1	(-)0,004
chi2 moyen					4,86					0,67

LA MÉMOIRE DES TERMES : ANALYSE CONCEPTUELLE

Manuel Célio CONCEIÇÃO¹

UCEH - Universidade do Algarve, Faro, Portugal

L' intérêt manifesté par les terminologues à l'égard des concepts et de la conceptologie a bouleversé les méthodologies de travail traditionnelles en terminologie.

Les termes peuvent être comparés à une pièce de monnaie, ayant deux faces inséparables : la face linguistique et la face cognitive. La face linguistique est un ensemble de morphèmes porteurs de signification qui servent d'étiquette au concept (face cognitive). Celui-ci est une unité d'un système, fonctionnant au niveau cognitif, et constituée par un ensemble de traits conceptuels qui le distinguent de tous les autres dans une conceptologie. Le concept n'est pas le signifié², puisque ce dernier existe au niveau linguistique; il est la manière dont on regarde le concept.

Dans le cadre des recherches en linguistique, l'analyse des dénominations (la face visible des termes) a été l'objet de la plupart des études terminologiques; les études morphosyntaxiques sont les plus fréquentes. Reconnue depuis toujours, et souvent citée, l'autre face du terme n'a commencé à être dépouillée avec soin que depuis quelques années. Il en va de même pour les relations entre les niveaux linguistique et cognitif en terminologie.

Les traits conceptuels qui caractérisent un concept sont «traduisibles» en éléments définitionnels qui constituent, à leur tour, le schéma définitionnel. Ce schéma, qui montre les relations entre les traits constitutifs de ce concept est «l'élément médian entre le terme et le concept» (Thoiron 1995b : 1). À partir de ce schéma définitionnel sont rédigées les définitions terminologiques, ou même encyclopédiques, et peut alors être créée la forme linguistique qui désignera le concept. La formation du mot³ se fera par le choix d'éléments de dénomination qui correspondent à certains (idéalement, tous) éléments définitionnels décrits dans le schéma. Quand les éléments définitionnels sont transposés au niveau linguistique, ils peuvent correspondre à un ou à plusieurs morphèmes et donc un ou plusieurs traits de signification.

¹ Mes remerciements aux professeurs Maria Teresa Lino (Universidade Nova de Lisboa) et Philippe Thoiron (Université Lumière-Lyon II) pour toutes les critiques et suggestions apportées.

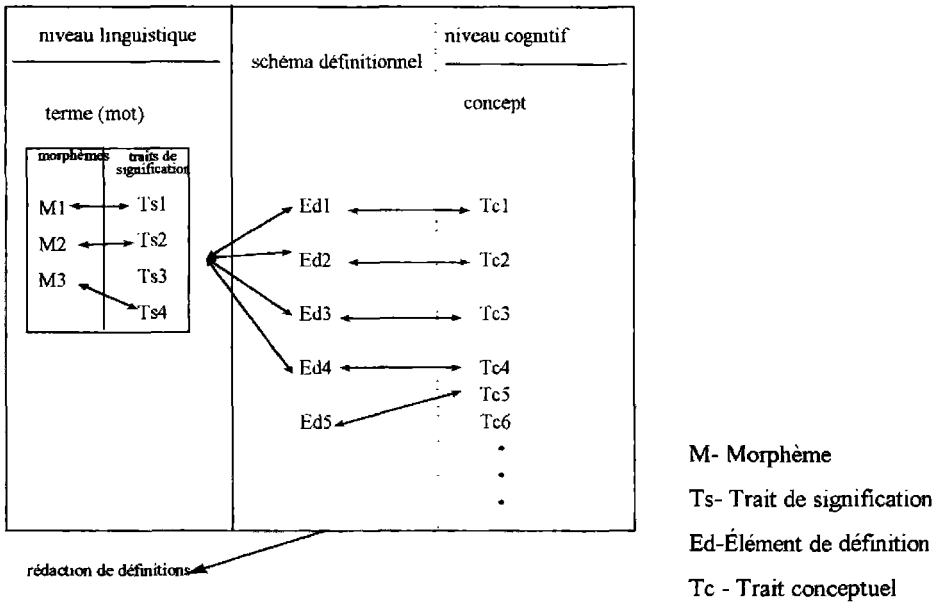
² Pour la distinction concept/signifié voir Thoiron (1995a : 118).

³ Nous entendons ici par *mot* le signe linguistique constitué par une séquence de caractères autorisée dans un système linguistique donné.

Le terme (forme linguistique) est donc un abrègement de forme et de signifié du schéma définitionnel. L'abrègement est, d'ailleurs, l'objectif de la nomination, et il peut provoquer, parfois, la démotivation et, donc, l'opacification.

Le terme (mot) est, ainsi, une synthèse linguistique du concept auquel on accède par l'analyse de ses comportements discursifs et par l'analyse morphosyntaxique de cette dénomination. Cette démarche se fait en associant aux formes des éléments de définition et, par voie de conséquence, des traits conceptuels, ou en les inférant à partir de l'ensemble des connaissances préalables. L'activation des traits conceptuels dépend de la combinaison des données linguistiques et textuelles. Les définitions rédigées seront d'autant plus complètes qu'elles seront fidèles au schéma et qu'elles rendront compte de tous les traits.

Chaque unité terminologique est donc la synthèse de la représentation présentée ci-dessous, adaptée de Thoiron (1995a).



Puisqu'une terminologie est une structure constituée des termes d'un domaine du savoir ou d'une sphère d'activité socioprofessionnelle et que la répartition de la réalité et du monde en domaines est subjective et socioculturellement déterminée, les termes sont le reflet d'une culture. Nous entendons ici culture au sens le plus large, en tant que processus dynamique dont les individus se servent.

Les concepts sont des unités de connaissance et ne sont donc pas fixes, ils sont relatifs et leur absolue universalité synchronique est discutable⁴. «Au plan sémantique [les concepts] sont des sémies stabilisées par les normes d'une discipline, et déliées de leurs

⁴ «The word "concept" is often taken as indicating a single fixed, well-delimited entity belonging to the intellectual sphere» (Langacker 1997). N'étant pas en accord avec ceci, l'auteur préfère *conceptualisations* à *conceptions*.

déterminations contextuelles» (Rastier *et alii* 1994 : 75). Les unités de connaissance sont situationnelles et relationnelles, par conséquent les progrès scientifiques et techniques ainsi que l'affinement des connaissances établissent de nouveaux traits conceptuels qui précisent les concepts. La moindre modification au niveau conceptuel entraîne forcément le changement du terme et l'évolution terminologique doit en rendre compte.

À la suite de ce qui vient d'être dit, nous nous sommes proposé d'analyser l'évolution terminologique/conceptuelle de quelques termes de pharmacologie en portugais. Sans vouloir faire des recherches diachroniques, l'objectif est uniquement de voir comment certains termes fondamentaux de ce domaine se sont «fixés» et de voir les liaisons entre traits conceptuels et éléments de nomination.

Les termes et leurs informations respectives ont été extraits de livres de pharmacologie (surtout pharmacopées) publiés depuis le XVIII^e siècle. Il faut, évidemment, analyser les textes sous l'angle de l'ordre herméneutique de la description, c'est-à-dire qu'il faut tenir compte des «différences de situation historique et culturelle qui séparent la production et l'interprétation» (Rastier *et alii* 1994 : 17). Il est indispensable de partir des textes, puisqu'ils sont un produit linguistique de l'interaction sociale dans laquelle les connaissances sont légitimées et validées.

Le XVIII^e siècle est dans ce domaine, comme dans beaucoup d'autres, le siècle par excellence du début de l'explosion terminologique. Les grandes découvertes dans les domaines connexes comme la chimie (rappelons Lavoisier) ont permis que de nouveaux termes enrichissent les pharmacopées héritées des grecs et des latins.

Contrairement au XVII^e siècle, caractérisé par une piètre fécondité terminologique, les grandes découvertes et les classifications destinées à mettre de l'ordre dans le naturel d'une part, l'expérimentation de plus en plus rigoureuse et ingénieuse et les progrès de tout ordre associés aux perfectionnements des techniques et des technologies d'autre part ont contribué à cette explosion terminologique. Les synonymies se sont multipliées et l'éponymie a été un phénomène très fréquent pour immortaliser les découvreurs.

Au long des siècles on peut constater une certaine pérennité dans le vocabulaire, mais on trouve également quelques modismes éphémères déterminés par toute sorte de changements, sociaux, scientifiques ou techniques.

Bien qu'il y ait fort peu d'études diachroniques en terminologie, et ce domaine ne fait pas exception, l'observation du surgissement ou de la disparition de certaines formes est tâche facile par rapport aux études sur le changement conceptuel. Ceci malgré la très grande difficulté de datation des formes (exception faite des formes éponymiques ou de certaines métaphores très stéréotypées).

Nous sommes, en outre, parti du XVIII^e siècle, parce que les premières pharmacopées écrites en langue portugaise datent de ce siècle — la première est de 1704 — et les premiers manuels de pharmacie (le terme pharmacologie n'est pas encore utilisé) datent de la fin de ce siècle.

Seuls les textes du XX^e siècle ont pu être numérisés, pour les autres la saisie a été faite manuellement. Pour chaque siècle nous avons pris un document officiel. Tous les textes ont été écrits en portugais par des spécialistes portugais et leur public est aussi spécialiste du domaine. Nous évitons ainsi les dérapages de sens liés à la vulgarisation. L'ensemble a été soumis à une lecture descriptive qui «résulte d'une description sémantique qui tient compte du sens immanent à la situation dans laquelle le texte a été produit» (Rastier *et alii* 1994 : 16).

Au début, nous avons l'intention de faire, non seulement une étude monolingue, mais aussi une analyse contrastive, prenant des textes français avec des caractéristiques éditologiques semblables. Cet objectif a été écarté au fur et à mesure de la comparaison des textes. En effet les différences ne sont pas pertinentes, étant donné que les textes sont très proches. Les textes du XVIII^e siècle ne sont souvent que des traductions et des adaptations des mêmes textes sources écrits en latin. À partir du XIX^e siècle, les textes portugais semblent s'inspirer beaucoup des textes français, ce qui témoigne de l'influence des recherches faites en France; influence reconnue par les auteurs qui ont esquissé l'histoire de la pharmacie et de la pharmacologie au Portugal.

Avec les termes ont été saisis les contextes et, si possible, les définitions présentées par les auteurs (cf. corpus de textes en annexe), vu que c'est dans le discours et dans la communication que s'opère la signification. Elle est le résultat d'une relation entre signifiés établie par la construction faite pendant l'interprétation.

Dans la construction de la signification, outre les éléments textuels considérés, il faudrait aussi voir les formules de reprise comme par exemple les procédés anaphoriques qui montrent les caractéristiques discontinues des traits significatifs (sèmes).

À partir de ces éléments textuels, l'objectif est de partir des désignations et des schémas définitionnels pour arriver aux traits conceptuels et aux concepts respectifs. Cette démarche peut permettre l'établissement d'une conceptologie.

Il s'agit donc de connaître les termes pour maîtriser les concepts. La maîtrise des concepts ne dépend pas seulement de la reconnaissance et de la compréhension des formants, d'autant plus qu'un concept peut avoir plusieurs désignations. L'arrivée au concept se complique aussi car sa connaissance n'est pas exclusivement linguistique, elle est aussi référentielle et propre aux pratiques sociales. C'est ce dernier aspect qui permet l'évolution conceptuelle associée aux événements relevant de l'histoire des sciences et des techniques.

La spécification des concepts peut se faire à partir de l'analyse du comportement discursif. Pour illustrer ce qui a été dit, voyons, à titre d'exemple, l'évolution du concept de médicament, à travers les informations extraites du corpus.

Médicament	
traits conceptuels	
1. forme pharmaceutique 2. produit altérations dans l'organisme 2.1 guérit 2.2 prévient 2.3 diagnostique 3. appliquée à l'homme et aux animaux	
définitions	traits présents dans la définition
« <i>medicamentum est quod naturam alterare potest; vel : medicamentum est illud quod corpori humano applicatum afflictiones nostras sanat; vel : medicamentum est quod instrumentum, quod medici operantur ad extirpandos; e sanandus morbis.</i> » (Santo António 1704 : 1) ⁵	2. « <i>naturam alterare potest</i> » 2.1 « <i>afflictiones nostras sanat</i> » ⁶ ? <i>instrumentum</i>
« <i>he tudo o que tomado por dentro ou aplicado por fora produz alterações em nossos humores</i> »(Vigier 1716 : 2)	2. « <i>produz alterações</i> »
« <i>agente terapêutico [...], substância ou substâncias, tiradas de algum dos três reinos da natureza, não essencialmente alimentares, mas que reduzidas a forma conveniente de administração, são capazes de modificar os actos do organismo, tornando-se assim úteis ao tratamento das doenças</i> » (Gomes 1815 : 2-3)	2. « <i>agente terapêutico</i> » ~1. « <i>forma conveniente de administração</i> » 2. « <i>modificar actos do organismo</i> » 2.1. « <i>úteis ao tratamento de doenças</i> »
« <i>toda a substância ou composição que possua propriedades curativas ou preventivas das doenças e dos seus sintomas, do homem ou do animal, com vista a estabelecer um diagnóstico médico ou a restaurar, corrigir ou modificar as suas funções orgânicas.</i> » (Pita, 1993 : 57, d'après le Décret n° 72/91 du 8 février)	~1 « <i>substância ou composição</i> » 2.1 « <i>curativas</i> » 2.2 « <i>preventivas das doenças e dos seus sintomas</i> » 3. « <i>do homem ou do animal</i> » 2.3 « <i>estabelecer diagnóstico</i> » 2.1 « <i>restaurar</i> » 2. « <i>modificar [...] funções orgânicas</i> »

Nous constatons que ce n'est qu'au XIX^e siècle que la nature de médicament se précise. Il commence par être «quod», puis, «tudo», «être ou substance», «agente thérapeutique», et devient «forma conveniente de administração». Il est curieux de constater qu'au XVII^e siècle ce concept incluait encore les instruments utilisés par les médecins, «quod medici operantur ad extirpandos, e sanandos morbis».

⁵ Reprend totalement la définition proposée par Galien (liv.I, chap.I).

⁶ Ce trait est aussi repris par Plenck (1786 : 2) «être ou substance qui fait cesser, par sa vertu médicale, la maladie».

La précision et le besoin de transformer la substance en composition et sous la forme d'administration vont de pair avec les changements qui se sont opérés lors du passage de la pharmacie galénique à la pharmacie chirurgicale, déjà présentée par Plenck (1786 : 9) comme «l'art de préparer ou de composer des médicaments».

L'art deviendra pharmacie chimique et industrie pharmaceutique⁷. Les médicaments ont cessé d'être simplement «uma amálgama de drogas onde as de origem vegetal prevaleciam» (Pita 1993 : 205).

Les transformations d'ordre scientifique, technique, et, par conséquent, conceptuel et même social justifient aujourd'hui l'existence de trois dénominations pour chaque médicament : dénomination scientifique, dénomination commerciale et dénomination commune.

Quoiqu'il en soit, le trait conceptuel que l'on infère à partir de la plupart de ces désignations et de ce terme (surtout) est traitement, associé au pouvoir de provoquer des altérations. Ce trait était déjà présent dans l'étymon latin MEDICOR (soigner, traiter) + MENT. Ce terme semble donc avoir une mémoire, même si sa référence a beaucoup changé⁸.

L'analyse conceptuelle facilitera l'identification du concept et sa distinction par rapport à ceux qui lui sont proches. La rédaction des définitions sera aussi facilitée. Dans ce cas, il faudrait analyser les concepts de remède⁹ et de drogue¹⁰, puisqu'ils sont souvent confondus avec médicament¹¹. Un autre terme parfois considéré à tort comme synonyme

7 «O grande comércio de drogas do Renascimento que se viu continuado pelo séc. XVII deu lugar ao grande comércio das especialidades farmacêuticas, fruto da laboração das gigantescas indústrias muitas vezes não apenas destinadas ao medicamento mas igualmente a outras finalidades, como por exemplo as indústrias químicas.» Pita (1993: 205) [Le grand commerce de drogues de la Renaissance continué au XVII^e siècle a cédé la place au grand commerce des spécialités pharmaceutiques, fruit du travail de gigantesques industries souvent destinées non seulement au médicament mais aussi à d'autres finalités, comme par exemple les industries chimiques.]

8 «Curative properties have been attributed to nearly every substance when it was new, unusual, or difficult to obtain and which can be forced into the human system. When the novelty of these substances wear off they cease to be used as medicaments». (Haggard 1946 : 388). Comme exemple, citons les pommes de terre qui, lors de leur arrivée en Europe, ont été considérées comme un médicament aphrodisiaque.

9 Remède (du latin Res Medeor) est un hyperonyme de médicament, ce n'est pas obligatoirement une forme pharmaceutique, c'est l'ensemble des moyens utilisés avec l'objectif de prévenir les maladies.

10 Drogue est un terme associé qui désigne la substance utilisée pour la préparation des formes pharmaceutiques. L'utilisation de ce terme est évitée à cause de la connotation sociale actuelle.

11 Cf. Vigier (1716 : 1-2).

«Pharmácia, que vem do nome grego Pharmacon, quer dizer medicamento, he a arte que ensina a preparar os remédios. [...] A Galénica he aquella que se contenta da simplez mistura, sem especulação das substâncias que cada hua das drogas he naturalmente composta». Plenck (1786 : 2). [Pharmacie, qui vient du grec Pharmacon, veut dire médicament, est l'art qui enseigne à préparer les remèdes. [...] La Galénique est celle qui s'occupe du simple mélange, sans spéculation des substances dont chaque drogue est naturellement composée.]

Cette confusion existe encore et beaucoup d'auteurs rappellent le besoin de correction terminologique (Pita 1993 : 205).

est «présentation pharmaceutique»¹².

Par la mise à jour systématique des traits sémantiques et conceptuels, on pourra voir les organisations conceptuelles successives et l'évolution des connaissances qui imposent une nature dynamique à la langue de spécialité¹³. Le stockage des résultats de cette analyse était probablement compliqué dans les bases de données terminologiques (BDT), il ne l'est plus dans les bases de connaissances terminologiques (BCT) dans lesquelles la distinction entre données linguistiques et données conceptuelles peut être faite.

L'analyse conceptuelle/linguistique des termes montre que certains traits se maintiennent, quelques-uns disparaissent et que d'autres s'ajoutent au fur et à mesure que les connaissances du domaine évoluent. D'ailleurs, on constate que les traits fondateurs accompagnent le terme, lorsqu'il devient formant d'un terme composé comme «preparação medicamentosa» [préparation médicamenteuse] ou «medicamento essencial» [médicament essentiel].

Du fait que «la scienza [è] come un gioco a tre : una mente che indaga, la natura chi responde e la comunita che debate» (Pera 1991), la mémoire du terme est mise en cause à chaque fois qu'il est utilisé et elle ne résiste que parce qu'elle constitue en quelque sorte son code génétique.

RÉFÉRENCES

- AAVV (1994) : *Studi sul trattamento linguistico dell'informazione scientifica*, a cura di Tullio di Mauro, Roma, Bulzoni Editore, 342 pages.
- CONCEIÇÃO, M. C. (1996) : «Terminologias em Português : uma questão de sobrevivência», M^a. H. Mateus (dir.), *Revista Internacional de Língua Portuguesa*, n° 15, Lisboa, Associação das Universidades de Língua Portuguesa, pp. 25-29.
- CONCEIÇÃO, M. C. (1996) : «Socioterminologia», *The Linguistic Construction of Social and Personal Identity. Special Conference on Sociolinguistics*, Évora, International Sociological Association - CIDEHUS, pp. 1-6.
- HAGGARD, H. (1946) : *Devils, drugs and doctors, the story of Healing from Medicine-Man to Doctor*, New York, Pocket Books, 490 pages.
- HARRIS, Z. (1988) : *Language and information*, New York, Columbia University Press, 170 pages.
- JARRONSON, B. (1992) : *Invitation à la philosophie des sciences*, coll. «Point Sciences», Paris, Seuil, 238 pages.
- LANGACKER, R. (1997) : «The contextual basis of cognitive semantics», J. Nuyts et E. Peterson (eds), *Language and conceptualization*, coll. «Language, Culture & Cognition», Cambridge, Cambridge University Press, pp. 229-252.
- LINO, M^a T., R. COSTA et M. CONCEIÇÃO (1996) : «Terminologia, informática e multimédia», *Jornada Panllatina de Terminologia, Perspectives i camps d'aplicació*, IV^e Réunion du Réseau Panlatin de Terminologie, Barcelone, IULA, Universitat Pompeu Fabra, pp. 73-81.

¹² Présentation pharmaceutique désigne le produit tel qu'il est délivré au patient ou au client de la pharmacie.

¹³ «L'énonciation et la compréhension de textes participent comme tous les échanges sémiotiques de l'évolution des cultures, et par là de la phylogénèse dont elle est l'aboutissement» (Rastier *et alii* 1994 : 22).

- OTMAN, G. (1996) : *Les représentations sémantiques en terminologie*, coll. «Sciences cognitives», Paris / Milan / Barcelone, Masson, 216 pages.
- PERA, M. (1991) : *Scienza e retorica*, coll. «Biblioteca di Cultura Moderna Laterza», Bari, Laterza, 268 pages.
- RASTIER, F., M. CAVAZZA et A. ABILLÉ (1994) : *Sémantique pour l'analyse, de la linguistique à l'informatique*, coll. «Sciences cognitives», Paris / Milan / Barcelone, Masson, 240 pages.
- ROBINSON, E. (1997) : «The cognitive foundations of pragmatic principles : implications for theories of linguistic and cognitive representation», J. Nuyts et E. Peterson (eds), *Language and conceptualization*, coll. «Language, Culture & Cognition», Cambridge, Cambridge University Press, pp. 253-271.
- THOIRON, P. (1995a) : «La terminologie multilingue et le traducteur», A. Hermans (dir), *Les dictionnaires spécialisés et l'Analyse de la Valeur*, coll. «Bibliothèque des Cahiers de l'Institut de Linguistique de Louvain», Louvain-la-Neuve, Peeters, 1997, pp. 117-130.
- THOIRON, P. (1995b) : «Concepts, termes et traduction», Communication présentée au 120^e Congrès National des Sociétés Historiques et Scientifiques, 23-29 Octobre 1995, Aix-en-Provence, 8 pages.
- WEISSENHOFER, P. (1995) : *Conceptology in terminology theory, semantics and word-formation*, «ITF-Series», Vienne, TermNet, XI + 270 pages.

ANNEXE

AAVV (1987) : *Farmacopeia Portuguesa V*, Lisboa, INCM.¹⁴

GOMES, António Bernardino (1815) : *Elementos de pharmacologia geral ou Princípios geraes de matéria médica e de terapêutica*, Lisboa, Typographia da Academia Real das Sciencias.

PINTO, Agostinho A.S. (1835) : *Pharmacographia do código farmacêutico lusitano ou Tratado de Pharmacodinamia*, Coimbra, imp da Universidade.¹⁵

PITA, João Rui (1993) : *Farmácia e medicamento, noções gerais*, Coimbra, Minerva.¹⁶

PLENCK, M. (1786) : *Pharmacologie médicale*, Paris, s/i.

SANTO ANTÓNIO, D. Caetano (1704) : *Pharmacopea Lusitana, methodo práctico de preparar & compor medicamentos na forma Galénica com todas as receitas mais uzuais*, Coimbra, Joam Antunes.¹⁷

TAVARES, Francisco (1794) : *Pharmacopeia geral para o reino e domínios de Portugal*, Lisboa, s/i.¹⁸

VIGIER, Joam (1716) : *Pharmacopea ulyssiponense, galénica e chymica que contem ...*, Lisboa, Pascoal da Sylva.

¹⁴ Texte officiel en vigueur et qui est une adaptation de *Pharmacopée Européenne*.

¹⁵ Deuxième pharmacopée officielle.

¹⁶ En ce qui concerne les définitions, cet auteur suit la loi en vigueur. Il cite souvent la Pharmacopée Portugaise et la législation où les concepts sont définis.

¹⁷ Cet auteur copie beaucoup d'extraits de textes en latin et les traduit. La définition de médicament qu'il présente, par exemple, est celle de Galien.

¹⁸ Première pharmacopée officielle.

DES MOTS QUI ONT LA MÉMOIRE LONGUE : LE GENOU, LA SOURIS ET L'ARBRE

Étienne TIFFOU

Université de Montréal, Canada

INTRODUCTION

La mémoire des mots, thème retenu pour ce congrès, se prête à différentes interprétations. S'agit-il des processus selon lesquels les mots s'ancrent plus ou moins durablement dans la mémoire des sujets parlants ? Ces processus sont d'ailleurs multiples. Il peut, en effet, s'agir de clarté de lecture d'un terme, d'associations insolites d'idées qui maintiennent sa vitalité, de l'aisance à en manier le paradigme, etc. Il peut également s'agir de l'art d'ériger un mémorial des mots, ce qui est une façon de leur assurer une certaine pérennité. Cet art intéresse, entre autres, mais pas exclusivement, la conception et l'élaboration de dictionnaires. Il n'est pas interdit non plus de prêter aux mots une mémoire grâce à laquelle il peuvent non seulement raconter des histoires perdues, mais aussi se survivre à eux-mêmes ou se régénérer.

De toutes ces possibilités, c'est la dernière que j'ai retenue. C'est donc une étude diachronique que j'entends mener. Mais celle-ci ne se limitera pas à une simple histoire de mots. Après avoir rappelé celle du genou, celle de la souris permettra de comprendre l'importance de la métaphore dans la création des mots; celle de l'arbre confirmera cette importance en montrant que la structure cognitive l'emporte sur une symbolique en principe d'autant plus contraignante qu'elle est inconsciente. Un tel résultat devrait nous inviter à nous interroger sur l'arbitrarité du signe linguistique pour conclure que celle-ci mérite d'être partiellement reconsidérée.

LE GENOU

On a fait remonter le genou à la même racine que *naître* et *connaître*. Selon cette hypothèse, on partirait d'une racine indo-européenne $*g^{(e/o)}n-$ suffixée soit en $*(e/o)H_1$, soit en $*(e/o)H_3$. Les radicaux comportant le premier suffixe seraient concernés par l'idée de naissance et ceux qui comportent le second seraient intéressés par l'idée de connaissance. Ainsi sur le thème I de la racine suffixée en $*(e/o)H_1$ ($*g^{e/on}-H_1$) ont été dérivés le grec *genevtrw*, le latin *genitor* et la sanskrit *janitá*, alors que sur le thème III ($*gn-H_1$) on a tiré les verbes givgnomai, (*g*)*nascor*, *jajñe*. Tous ces termes sont sémantiquement concernés par l'idée de naissance. La racine suffixée en $*(e/o)H_3$ est attestée surtout au thème II (ex. : *gignwvskw*, *co-gnósko*, *jñá-*) et réfère toujours à l'idée de connaissance. Quant au

thème du genou, il est très probable qu'il remonte à la même racine, sans qu'on puisse dire si ce thème était suffixé ou pas. Dans le cas où il le serait, nous aurions affaire à un thème I sans que l'on puisse postuler en toute certitude l'une des deux laryngales, puisque celles-ci tombent devant voyelles. Si l'on admet qu'il faille poser une développante, il est fort probable qu'il s'agisse plutôt d'un *H₁*. Toutefois la forme sanskrite *jánu* donne à penser que ce terme n'était pas suffixé. En effet, la longue peut s'expliquer par la loi de Brugmann; or, si le *-n-* était suivi d'une laryngale, la voyelle indo-européenne aurait été en syllabe fermée et la loi n'aurait pu de ce fait s'appliquer. Au reste, le thème pour désigner le genou est fort stable dans les langues indo-européennes, quoiqu'il présente selon celles-ci une alternance de timbre : *govnu* en grec, *genú* en latin et en hittite *genu*.

Toutefois, cette parenté ne peut être absolument prouvée. P. Chantraine note bien dans son dictionnaire étymologique de la langue grecque (1968 : 233) qu'on s'est demandé «si les deux familles divergentes de *gignwvskw* et de *givgnomai* ne devraient pas être rapportées au nom du genou.» Malheureusement, cette hypothèse, ajoute-t-il quelques lignes plus bas, ne peut se démontrer rigoureusement et il n'est guère possible de faire de *givgnomai* et *gignwvskw* des dénominatifs du nom du genou. Quant à Ernout et Meillet (1951 : 273B), ils penchent en faveur de ce rapprochement, car, selon eux, il y a suffisamment d'éléments pour rendre cette étymologie tout à fait vraisemblable. D'ailleurs, pour que preuve soit faite, pourquoi est-il nécessaire que *givgnomai* et *gignwvskw* soient des dénominatifs du nom désignant le genou ? En outre, les remarques de Meillet (1926 : 54-55) sur l'adjectif anglais *genuine* évidemment dérivé sur celui-ci, plaide en faveur d'un rapprochement avec l'idée de naissance et de connaissance. Au reste, la forme latine invite à se demander si son vocalisme ne serait pas dû à une influence de *geno*. Enfin et surtout, d'anciens textes montrent à l'évidence que le genou jouait une place importante dans la reconnaissance du nouveau-né, ce qui le relie clairement à l'idée de naissance et de connaissance.

Le survol des mots dérivés de la racine **g^(e/o)n-* ne laisse pas, en effet, d'intriguer. Quels rapports de sens peut-il y avoir entre la naissance, la connaissance et le genou ? En français, les deux premiers termes sont clairement apparentés; ce rapport formel maintient de façon fossilisée une relation entre ceux-ci. L'hébreu présente, au demeurant, une affinité similaire, mais il semble bien que les idées de naissance et de connaissance soient reliées l'une à l'autre de façon différente. La connaissance précède la naissance, puisqu'il fallait qu'Adam connût Ève pour que Caïn et Abel en naquissent. D'après de vieilles légendes, il ressort qu'il en va au rebours en indo-européen et que ce soit la naissance qui précède la connaissance. Cela s'impose si l'on veut intégrer le genou dans le jeu des relations sémantiques. Dans une version hittite éditée par Gaster (1953 : 101-122) d'un conte hourrite, remontant fort probablement lui-même à une version sumérienne plus ancienne, Koumarbi féconde une montagne qui enfante «un énorme galet en forme de nouveau-né» :

«...les déesses s'avancèrent pour le recevoir, le portèrent tendrement à Koumarbi et le placèrent sur ses genoux.

Koumarbi était transporté de joie; prenant l'enfant dans ses bras, il l'éleva en l'air et ensuite le berça sur ses genoux en chantonnant [...] il cessa enfin, installa l'enfant commodément sur ses genoux : "Petit, dit-il, il me faut te donner un nom; tu t'appelleras Oullikoumi"» (1953 : 106-107).

Ce texte revient à plusieurs reprises sur l'installation du nouveau-né sur les genoux de celui qui l'a engendré. À ce propos, Gaster (1953 : 117) note un peu plus loin :

«...l'enfant est placé sur les genoux de son père; c'était là, dans les temps anciens, une façon habituelle de reconnaître la paternité, les genoux étant alors supposés avoir un rapport anatomique direct avec la génération; et l'adjectif anglais *genuine* (véritable, authentique) signifie à l'origine "qui a été mis sur les genoux de son père pour être reconnu"».

Plutôt que de voir dans le genou «un rapport anatomique direct avec la génération», il est préférable de le considérer comme l'organe de l'engendrement juridique. Cette symbolique est sans nul doute bien connue des indo-européens. La naissance de Dionysos qui sort de la cuisse de Zeus après qu'il a été récupéré du sein de sa mère Sémélé foudroyée par la vision du roi des dieux dans sa toute puissance, est une transposition de la reconnaissance juridique de l'enfant qui vient de naître.

Au terme de cette brève présentation, le mot genou ouvre le champ à la mémoire sur des pratiques biologiques et juridiques. L'origine qu'il partage avec le mot *gonade* s'explique ainsi. Les mots portent en eux les clés qui donnent l'accès à la connaissance de temps révolus dont la documentation est pratiquement inexistante. Le cas du genou est intéressant; il diffère, en effet, des deux autres cas que nous allons examiner, car il est contextuel et ne peut être compris sans référence à l'idée de connaissance et de naissance. Cet état complexe sans éradiquer la métaphore sous-jacente l'a déstabilisée. Ce terme porte témoignage à un type de mémoire à moitié effacée à laquelle il assure néanmoins une certaine pérennité.

LA SOURIS

La souris se prête à bien des métaphores relativement récentes qui n'intéressent, de ce fait, que médiocrement notre propos. Ainsi la désignation familière de femme sous le nom de *souris* n'est pas riche d'enseignement. On appelait *souris de rempart* les personnes du beau sexe trafiquant de leurs charmes avec les militaires auprès desquels elles se glissaient avec la discrétion furtive d'une souris. Amputé de son déterminant, ce terme devait par la suite s'appliquer à celles qu'on appelait également à cette époque des lorettes; la connotation péjorative qu'il comportait s'est finalement perdue et ce mot désormais s'emploie familièrement pour parler d'une jeune femme (cf. Robert : 1996B). En revanche, confronté à un mot latin différent, mais peu éloigné sémantiquement, le nom de la souris se prête à des analyses beaucoup plus révélatrices qui permettent de raviver de vieux souvenirs.

La souris vient d'un mot latin *sorex*, *soricis* attesté dans les *Res Rusticae* de Varron. Ce terme est indo-européen et serait dérivé d'une racine suffixée **s^(e/o)w-(^{e/o})r-*. Le mot latin serait tiré du thème II (**sw-er- > *swor- > sor-*), alors que son correspondant grec *u[ra]*x viendrait du thème III (**sw-r- > hur-*). Il est parallèle à un nom racine indo-européen attesté notamment en latin (*mús*), en grec (*mu~*) et en sanskrit (*múl-*); selon Chantraine (1984 : 725B), la longue s'expliquerait par le caractère monosyllabique de ce mot. Celui-ci a survécu en français dans le nom de la musaraigne «la souris/le rat araignée» qui passait pour être venimeuse. C'est la seule attestation que nous ayons de ce

terme avec sa valeur propre. Il est mieux connu dans un autre sens dérivé par métaphore, mais celle-ci n'est plus perçue des sujets parlants. Il s'agit du muscle.

À vrai dire cette métaphore était largement répandue dans bien des langues indo-européennes. En s'appuyant sur l'analogie de forme, voire de mouvement, des biceps, on a désigné le muscle comme le «petit rat» ou la «petite souris»; ainsi le nom du rat en grec *mu~* désigne également le muscle. Quant au latin, il a adjoint au mot *mūs* un suffixe de diminutif à valeur expressive : *-culu(m)*. Ce suffixe est bien attesté en français où la plupart des mots en *-eil* remontent à ce suffixe : *sommeil* < *somniculu(m)*; *oreille* < *auriculu(m)*; *soleil* < *soliculu(m)*, etc.). Le muscle signifierait littéralement le «souriceau». Il n'y a pas lieu de distinguer la souris du rat; d'ailleurs l'opposition entre les termes rat et souris n'est pas claire et n'influence pas, à l'origine, la métaphore.

Il n'y aurait pas grand chose à ajouter sur l'étymologie du mot muscle, si la métaphore perdue n'était pas réapparue sous une autre forme. En effet, de nos jours, il est un petit muscle vers le haut du manche du gigot d'agneau que l'on nomme la souris. La métaphore est claire pour les sujets parlants et à ceux qui s'étonnent de cette appellation on explique que la forme de ce muscle fait penser à une petite souris. Nous aurions là un cas de résurrection métaphorique intéressante. Celle-ci n'est pas cependant aussi évidente. Le terme souris avait, en effet, au siècle dernier une signification moins restreinte :

«Par analogie de forme avec le corps de la souris ou un de ses membres, le mot s'est employé comme le latin *mus*, et surtout son dérivé *musculus* (→ muscle), désignant la partie charnue du bras, de la main ou de la jambe (v. 1250 *soris*), l'espace de la main entre le pouce et l'index (1690), sens disparu au XIX^e s.; *souris* est le nom encore usuel (1694) d'un muscle charnu à l'extrémité du gigot, contre l'os.» (Robert 1994 : 1996B)

Il ressort de cette citation que la souris n'avait pas, lorsqu'elle était employée métaphoriquement un sens aussi spécialisé que celui qu'il a de nos jours. Est-ce à dire que la métaphore a été recréée avec ce terme au cours de l'histoire du français ? Ou est-ce à dire que celui-ci avait parallèlement eu les mêmes emplois que *mus* qu'il avait fini par perdre au profit de ce dernier ? Il est certain que dans ce cas, l'histoire de ce mot est beaucoup moins riche d'enseignement et ne ferait que témoigner une fois de plus des combats que les mots se livrent entre eux pour assurer leur survie.

S'il est permis de laisser planer, faute de documentation, un doute sur la solution à retenir, un élément nouveau permet de redonner à ce type de métaphore l'intérêt qu'il mérite. Cette métaphore n'est pas propre à un groupe de langues apparentées, mais se retrouve dans d'autres familles linguistiques. Ainsi, en arabe (exemple donné par A. Mehiri, professeur émérite de l'Université de Tunis), *fār* désigne la souris mâle et *fāra* la souris femelle. Ce dernier mot est aussi employé pour parler du muscle de la jambe, alors que le premier peut également signifier la «chair» du dos. Le bourouchaski, langue isolée de l'Himalaya, présente également ce type de métaphore. Le mot pour rat, souris se dit *gīrkis*, or on lit dans le lexique de Berger (1974 : 145B) : *gīrkis cap* «Biceps» et Lorimer, dans son dictionnaire (1938 : 168A), propose pour le dialecte du Hounza le terme *pīni gīrkis* qu'il glose en «muscle of calf of leg. Cp. Latin *mus* : *musculum*». Il est sûr que les Bourouchos ont élaboré cette métaphore de façon indépendante. Cela montre à l'évidence que nous avons là une image de type universel qui peut s'imposer dans des

milieux totalement différents indépendamment de toute influence extérieure. Un tel type de métaphore fait appel en quelque sorte aussi bien à une mémoire dans le temps qu'à une mémoire dans l'espace.

L'ARBRE

Le nom de l'arbre est bien connu en indo-européen; il repose sur une racine suffixée $*d^{(e/o)}r-(e/o)w-$ dont les trois thèmes sont bien représentés; ex. : thème I : gr. *dovru*. skt. *dáru* ($*dor-w$); thème II : gr. *devndreon*, angl. *tree* ($*dr-ew$); thème III : gr. *dru~w* «le chêne» ($*dr-w$). Employé de façon métaphorique, il est devenu symbole de la stabilité et de la solidité aussi bien au sens propre qu'au sens figuré. C'est ainsi que nous devons à Hésychius d'avoir conservé le mot *droÛovn* qu'il glose en *ijscurov* «fort, solide». Les langues germaniques offrent de bons exemples du sens figuré, tels le *treu* de l'allemand «loyal, fidèle» et le *truth* de l'anglais «vérité».

Il est possible, selon moi, de remonter plus haut en voyant dans le nom de l'arbre une métaphore très ancienne. Cette hypothèse que j'ai conçue il y a plusieurs années, ne laissera pas d'étonner, mais j'ai été rassuré en trouvant dans le livre récent de Gamkrelidze et d'Ivanov (1995) une étymologie proche de celle que j'avais parallèlement élaborée. La racine $*d^{(e/o)}r-$ fait penser à celle que l'on a postulée pour la peau et que l'on retrouve dans le grec *devrma* «peau». Ce nom comporte un suffixe $*-m^{(e/o)n}$, suffixe de nom d'action permettant de dériver des déverbatifs. De fait, il dérive d'un thème verbal *devrw* attesté avec le sens d'«écorcher, dépouiller». On ne s'étonnera pas, dans ces conditions, que le mot *devrma* désigne la peau que l'on arrache. Ainsi que le note, Chantraine (1968 : 265B) «*devrma* n. a dû se dire d'abord de la peau dépouillée d'un animal [...] de peaux préparées pour faire des sacs, etc.» Le rapprochement formel entre les noms signifiant respectivement la peau et l'arbre ne manque pas de vraisemblance, mais cette ressemblance peut être fortuite et résulter de neutralisations. C'est un avatar qui se produit fréquemment et qui est bien connu des comparatistes. D'autre part, le rapport sémantique entre les deux mots est loin d'être clair. Ce n'est qu'en trouvant un lien acceptable entre eux que l'hypothèse pourra avoir un certain fondement.

Gamkrelidze et Ivanov (1995 : 202) remarquent qu'il est possible de comparer sémantiquement les thèmes de l'arbre avec celui de la peau, ou plus exactement de son arrachement :

«This base, meaning 'tree; oak' (c'est-à-dire $*t'e/oru-$ et $*t'reu-$) [...] can be semantically compared to **t'er-** (Lith. *derù, dirti* 'cut off'. Latv. *nuðdara* 'pole with branches cut off', Russ. dial. *dor* 'logs that split easily') and $*t'r-oH$ (Oicel. *troD(a)* 'pole', Slav. *drati* 'remove skin or bark'...). This may indicate [...] that the form is an extended root in Benveniste's sense with two ablaut states, $*t'or-u-/t'r-eu-$ »

Il ressort de cette citation que le radical **t'er-** s'applique à un élément en bois fragile ou émondé; mais c'est surtout, et c'est également notre hypothèse, l'idée d'enlever l'écorce qui s'impose. L'arbre serait en quelque sorte «l'écorcé» ou «l'écorçable». Cette idée est liée essentiellement, selon moi, à celle du chêne, et ce de deux façons. Il peut s'agir du chêne liège; lorsque le liège a été récolté, les forêts dont il provient semblent être formées d'arbres écorchés. On peut également proposer une explication moins restrictive; le bois du chêne est extrêmement solide et entre dans la fabrication de divers objets résistants (ex. : skt

druvāya «plat en bois»; *drún5am* «épée», *drun5í* «seau», etc.), ce qui suppose que ce bois pour être utilisé devait être le plus fréquemment écorcé. Au reste, ces deux points de vue ne s'excluent pas et ont pu concourir à motiver la désignation de l'arbre.

Si cette hypothèse mérite d'être retenue, il est clair que la métaphore remonte à une date très ancienne et qu'elle a été occultée par d'autres métaphores que nous avons évoquées et à la base desquelles se trouve l'arbre. Où donc est la mémoire de ce mot ? Dans l'émergence de la métaphore dans d'autres termes d'origine différente. C'est ainsi que le français a retrouvé un rapport entre l'arbre et la peau arrachée dans les termes «écorce» et «écorché». D'un point de vue synchronique la ressemblance est frappante et impose à l'esprit un rapport de sens évident. Malheureusement, l'origine de ces deux noms est différente. À propos de l'écorce on trouve dans le *Robert étymologique* (1994 : 654B) :

ÉCORCE n. f. représente l'aboutissement (1176) du latin impérial *scortea* «manteau de peau». Ce mot dérive de *scortum* «peau, cuir», qui se rattache à une racine indo-européenne *sker- ou *ker-...

Le mot apparaît dans le sens d'«enveloppe des troncs, des branches, d'arbres» qui correspond au latin *cortex*...

Quant à écorcher, on lit dans la rubrique qui lui est consacrée :

ÉCORCHER v. tr. est issu (1155) du bas latin *excorticare* «écorcer», composé de *ex-* et du latin classique *cortex, corticis* «écorce», qui se rattache à une racine indo-européenne *sker- «couper», «séparer», «partager». (1994 : 654B)

Ces étymologies ne sont divergentes qu'en apparence. Il est vrai qu'écorce vient de *scortum*, alors qu'écorcher dérive de *cortex*. Mais ces deux mots ont même origine et remontent, comme il est signalé dans le Robert, à un thème indo-européen *sker-. La métaphore s'était donc reconstituée à date plus ancienne que le français. Celui-ci met, au reste, en évidence la force de la métaphore, en détournant les termes latins de leur valeur propre; en effet, le terme qui s'applique à l'arbre dérive de celui qui signifie «peau», tandis que celui qui s'applique à la peau dérive de celui d'«écorce».

C'est ainsi que cette métaphore, apparemment insolite, se trouve fondée par un jeu de mémoire qui ne restitue pas un état ancien, mais réassigne à des termes nouveaux apparentés une vieille relation métaphorique plus ancienne que celle qui lie la notion d'arbre à celle de vérité.

INTERPRÉTATION

Cette présentation rapide de trois mots à la mémoire longue invite à les caractériser dans un premier temps les uns par rapport aux autres. En fait, le «genou» s'oppose clairement aux deux derniers. Il s'agit d'un cas classique de reconstitution historique à partir de données linguistiques. Là où la documentation en textes ou en vestiges manque, l'histoire des mots peut apporter de précieuses informations. É. Benveniste a donné dans ses *Études indo-européennes* (1969) une brillante démonstration de ce que cette méthode permettait d'obtenir. On ne pouvait pas ne pas rappeler dans une étude consacrée à la mémoire des mots ce genre de démarche.

La souris et l'arbre introduisent de façon complémentaire, et c'est pour cela que je les ai retenus, à des réflexions d'un ordre différent où synchronie et diachronie finissent par se conjoindre. On verra une des oppositions particulièrement éclairante dans leur valeur symbolique. La souris, en effet, en est très peu chargée. Dans leur *Dictionnaire des symboles* (1982 : 904B), Chevalier et Gheerbrant lui consacrent moins d'une demi colonne et encore n'est-ce que pour faire allusion à des croyances bambaras qui lient la souris à l'excision des jeunes filles. En revanche, ces deux auteurs consacrent à l'arbre treize colonnes (1982 : 62A-68A) tout au long desquelles ils présentent les mythes liés à l'arbre, «l'un des thèmes symboliques les plus riches et les plus répandus; celui également dont la bibliographie, à elle seule, formerait un livre» (1982 : 62A). Or il est significatif que les valeurs symboliques que l'on prête à l'arbre ne soient pas inscrites dans la métaphore qui vient d'être présentée. Celle-ci est de type cognitif et est totalement comparable à celle de la souris et du muscle qui repose sur une analogie de forme et peut-être de mouvement, et sur rien d'autre. Cela peut mettre en cause la lecture que l'on pourrait faire d'archétypes, dans le sens de Jung, à partir du langage. Les mots sur le divan, pour reprendre l'expression de Nouss, ont beaucoup moins à dire qu'on ne pense. On a plus à apprendre d'eux sur le fonctionnement de notre esprit que sur celui de notre inconscient. C'est du moins la première leçon que je suis tenté de tirer de cette brève, trop brève analyse.

Ce genre de métaphore que je qualifierais de type cognitif permet de tirer quelques informations sur le processus de dénomination. Tout d'abord ce processus tend à une certaine universalité. La liaison de l'arbre et de la peau est généralisée tout au long de l'histoire de l'indo-européen et de ses dialectes, puisqu'elle est sous-jacente dans le rapprochement à faire entre deux racines de sens apparemment très opposé, qu'elle continue de survivre, à vrai dire bien masquée, dans le *scortum* et le *cortex* des Latins et qu'on la retrouve enfin remise en lumière en français avec l'écorce et l'écorché. La souris est encore plus significative, car on la retrouve liée au muscle dans des langues qui ne sont pas d'origine indo-européenne. Ce caractère universel invite à considérer que la production lexicale n'est pas toujours de type strictement linguistique et que des schèmes mentaux qui en sont indépendants peuvent la conditionner. Sans prétendre trancher la question inépuisable des rapports de la pensée et du langage, nous avons ici un cas assez clair de préséance de la pensée sur le langage.

Enfin, à la lumière de ce qui précède, on peut rappeler quelques nuances à apporter à l'arbitrarité du signe linguistique. On a déjà mis en évidence la motivation intersigne; il est évident que les ressemblances que l'on décèle entre «jardin» et «jardinage» n'est pas le fruit d'une décision non motivée. Des mots, comme ceux que nous venons d'analyser, montrent à l'évidence qu'il peut y avoir également, dans certains cas, motivation entre le signifié et le référent. C'est dans les structures de notre esprit qu'il faut voir les raisons décisives de cette motivation. Le caractère universel de certaines métaphores en donne une amorce de preuve qui est loin d'être négligeable.

CONCLUSION

De même que l'on se rappelle les mots, les mots se souviennent. Ce souvenir est parfois crypté, mais, clair ou opaque, il est porté par une métaphore. C'est donc à l'étude de la métaphore que renvoie la mémoire des mots, ce que j'ai, au demeurant essayé de mener partiellement, il est vrai, au long de cet exposé. Lakoff et Johnson (1985) se sont attachés à étudier la place de la métaphore dans les situations quotidiennes de communication. D'après eux, il existe des principes selon lesquels les métaphores s'organisent. Ainsi une idée d'agression ou d'offensive est liée aux activités intellectuelles :

« Vos affirmations sont *indéfendables*. Il a *attaqué chaque point faible* de mon argumentation. Ses critiques visaient *droit au but*. J'ai *démoli* son argumentation. Je n'ai jamais *gagné* sur un point avec lui. Tu n'es pas d'accord ? Alors, *défends-toi* ! Si tu utilises cette *stratégie*, il va t'écraser. Les arguments qu'il m'a opposés ont tous *fait mouche*. » (1985 : 14)

L'approche de Lakoff et Johnson est très intéressante, mais le genre de métaphore étudiée est complexe, car toujours lié à un contexte, et éminemment variable selon le type de société. Mon point de vue, surtout historique, est peut-être plus fondamental, car les métaphores qui m'intéressent sont liées au mot qui les charrie indépendamment de tout contexte. Dans cette mesure, elles portent en elles une certaine universalité; mais avant que de poser les principes sous-jacents aux métaphores marquées de cette universalité, il est nécessaire d'en recueillir le plus grand nombre possible. J'en ai proposé un maigre bouquet; il faudrait travailler plus longtemps aux champs¹. D'autre part, la validité de la métaphore doit être testée diachroniquement; il faut, en d'autres termes, en décrire la mémoire qu'en fait la parole et voir si celle-ci ne la réajuste pas au goût du jour. Si tel est le cas, diachronie et synchronie s'harmonisent pour porter témoignage à la stabilité que l'esprit maintient au delà de ses changements, voire de ses errances. J'espère que la mémoire des trois mots que j'ai évoqués contribuera à montrer comment cette harmonie finit, dans certains cas, par être atteinte.

RÉFÉRENCES

- BENVENISTE, É. (1969) : *Le vocabulaire des institutions indo-européennes*, Paris, Les éditions de minuit, 2 vol.
- BERGER, H. (1974) : *Das Yasin Burushaski*, Wiesbaden, Otto Harrassowitz.
- BERT, M. et J. SERME (À paraître) : « A la recherche des tendances universelles dans l'évolution sémantique des termes désignant des parties de la tête ».
- BOISSON, C. (À paraître) : « Sur le trope lexicalisé 'la bouche du couteau' = le tranchant du couteau ».

¹ Au reste ce travail est en cours. Témoins l'article de Brown, C. H. et S. R. Witkowski (à paraître) ainsi que l'étude de C. Boisson (à paraître) et celle de Bert Michel et J. Serme (*A la recherche des tendances universelles dans l'évolution sémantique des termes* (à paraître)).

- BROWN, C. H. et S.R. WITKOWSKI (1981) : «Figurative language in a universalist perspective», *American Anthropologist*, 8, 3, pp. 596-615
- CHANTRAINE, P. (1968 et 1984) : *Dictionnaire étymologique de la langue grecque*, Paris, Klincksieck, 2 vol.
- CHEVALIER, J. et A. GHEERBRANT (1982) : *Dictionnaire des symboles*, Paris, Robert Laffont/Jupiter.
- Dictionnaire historique de la langue française* (1994) : Paris, Dictionnaires LE ROBERT, 2 vol., nouvelle édition.
- ERNOUT, A. et A. MEILLET (1951) : *Dictionnaire étymologique de la langue latine*, Paris, Klincksieck, 3^e édition.
- GAMKRELIDZE, V. et V. IVANOV (1995) : *Indo-european and the Indo-europeans. A reconstruction and historical analysis of a proto-language and a proto-culture*, Berlin, New york, Walter de Gruyter, 2 t.
- GASTER, T. (1953) : *Les plus anciens contes de l'humanité*, Paris, Payot.
- LAKOFF, G. et M. JOHNSON (1985) : *Les métaphores dans la vie quotidienne*, Paris, Les éditions de minuit, traduit de l'américain par M. Defornel avec la collaboration de J.-J. Lecercle.
- LORIMER, D.L.R. (1938) : *The Burushaski Language. Vocabularies and index*, Oslo, Instituttet forsammenlignende kulturforskning.
- MEILLET, A. (1926) : *BSL*, XXVII, p. 54 sq.

LA MÉMOIRE DU COMMERCE : ASPECTS ÉTYMOLOGIQUES ET PHRASÉOLOGIQUES

Jeanne DANCETTE et Christophe RÉTHORÉ

Université de Montréal, Canada

INTRODUCTION

En 1949, en France, Édouard Leclerc, fondateur de la chaîne de grandes surfaces du même nom, décide d'appeler ses magasins «centres de distribution». Dès 1950, les nouveaux commerces qui, comme Leclerc, pratiquent le libre service se désignent également sous le nom de «distributeurs», plutôt que «commerçants». L'habitude est prise et le commerce moderne s'appellera désormais «distribution» (Benoun et Héliès-Hassid 1995 : 15).

Ce changement de paradigme, de commerce à distribution, a intrigué les auteurs de cet article, qui travaillent depuis plusieurs années sur le vocabulaire de la distribution (Dancette et Réthoré 1997). Nous avons voulu en examiner les causes et les modalités. Nous sentions que ce changement de direction terminologique ne s'expliquait pas par une différenciation sémantique fondamentale, par des traits qui s'opposeraient tels que le commerce s'oppose à l'industrie, par exemple, mais plutôt par les connotations négatives attachées à l'ensemble {commerce, commerçant, commercial, etc.} et dont il fallait se défaire pour redorer le blason du commerce.

Notre étude de cette question nous a conduits à chercher une explication dans les développements de l'histoire économique à travers les âges. Ensuite, nous avons examiné comment les dictionnaires de français, du 16^e au 20^e siècle, attestaient de l'évolution de ces termes. Enfin, par l'étude lexicométrique de la presse actuelle française et québécoise, nous avons cherché à voir si l'on retrouvait dans la société d'aujourd'hui le jeu de connotations que nous supposions, généralement négatives pour l'ensemble {commerce, commerçant, etc.} et neutres ou positives pour {distribution, distributeur, etc.}, en analysant comment de telles connotations opéraient dans des cultures différentes (France et Québec).

1. HISTORIQUE

1.1 Des marchands du Temple au marché noir, en passant par la route des Indes

Depuis ses origines, le commerce suscite des sentiments contradictoires, de l'admiration au mépris. De tout temps et dès l'Antiquité, les grands peuples voyageurs et

marchands, les Crétois, les Phéniciens, les Vénitiens, font naître l'admiration (Attali 1991). Les grands exploits liés à la découverte de la route des épices et de la soie nourrissent l'imagination et donnent lieu à une riche littérature mettant à l'honneur la débrouillardise et l'intelligence du voyageur, ainsi que l'exotisme des contrées lointaines.

«Je me rendis à Bassora, où je m'embarquai avec plusieurs marchands sur un vaisseau que nous avions équipé à frais communs.

Nous mîmes à la voile, et prîmes la route des Indes orientales par le golfe Persique...»

Les Mille et Une Nuits. Histoire de Sindbad le Marin (1996 : 11)

Dans l'histoire des nations, le commerce est reconnu comme facteur indéniable d'évolution, mais il se heurte au conservatisme des forces en place. Les peuples marins, marchands, s'opposent aux peuples terriens, guerriers et paysans. C'est la thèse de Fernand Braudel (1985 : 179) : jusqu'en 146 avant J.-C., Carthage, convoitée et détruite par Rome, symbolise la lutte «d'un peuple essentiellement marin et marchand et d'un peuple essentiellement terrien, guerrier et paysan». Un peu plus tard, poursuit Braudel, l'essor du commerce en Islam fait apparaître le clivage entre une Méditerranée orientale développée, où l'on connaît la monnaie, la lettre de change et le chèque — en arabe *chakk* — et une Méditerranée occidentale, barbare et sous-développée. Les villes situées à l'articulation de ces deux mondes occupent une position pivot. À Carthage, comme à Bagdad ou à Damas, convergent les esclaves d'Afrique, le plomb, le cuir d'Andalousie et de Sardaigne, l'or de l'Afrique noire, le blé d'Afrique du Nord, etc.

«Je veux porter du safran persan à la Chine, où j'ai entendu dire qu'il y a grand prix, et ensuite de la porcelaine de Chine dans la Grèce, du brocard grec dans l'Inde, de l'ac indien à Alep, du verre d'Alep dans le Yémen et des étoffes rayées du Yémen en Perses ...» (Al-AQâsim al-Harîrî 1992)

D'une manière générale, l'essor du commerce entraîne la mise en place d'une économie monétaire, voire d'un capitalisme marchand où les fonctions de banquier et de commerçant tendent à se confondre. D'une part, le commerce est valorisé, car lié au développement des puissances; d'autre part, il fait l'objet d'une position ambiguë, y compris des souverains, qui encouragent ou freinent le commerce, selon les moments et les valeurs religieuses dominantes. Dans les pays d'Europe occidentale à forte tradition catholique, comme l'Italie, l'Espagne, le Portugal et la France, les activités commerciales sont généralement regardées avec mépris, à la différence des pays de religion protestante, et notamment calviniste, qui n'ont pas les mêmes réticences idéologiques face à l'argent (Charmes 1989 : 17-18; Weber 1964; Chung, Shepard et Dollinger 1989 : 309-311). L'histoire et l'origine des grandes fortunes à travers le monde montrent un lien évident entre les valeurs religieuses et culturelles d'un pays et l'enrichissement de ses ressortissants¹.

¹ L'influence de la religion sur le comportement économique des individus est documentée. Charmes (1989 : 16) donne les chiffres suivants qui datent des années 80. Aux États-Unis, sur un total de 65 milliardaires «recensés», on comptait un seul catholique et 64 protestants, dont 59 calvinistes. En France, parmi les 50 plus grosses fortunes, on trouvait 5 catholiques, 5 juifs et 40 protestants, dont 33 calvinistes. Il y a, selon ses termes, une «disproportion statistiquement significative» entre ces chiffres et la répartition religieuse que cet auteur illustre par le tableau ci-dessous (*op. cit.*, 16) :

Quelle qu'en soit l'origine, les nouvelles fortunes font naître l'envie et revivre l'image des marchands du Temple, agressifs et mercantiles. Toutefois, les clichés, que nous retrouvons jusqu'au 20^e siècle, affectent surtout les moins fortunés des commerçants, présentés comme étant près de leurs sous, voire malhonnêtes et insensibles.

Au cours de la Seconde Guerre mondiale, les événements et le climat qui règnent en France à cette époque (occupation, collaboration, marché noir, etc.) ne font rien pour aider le commerce à se bâtir une nouvelle réputation, ainsi que le montrent ces citations empruntées à divers auteurs.

«À cirer la botte allemande, les bénéfiques avaient été reluisants. Que l'on se donne la peine de consulter les chiffres : 10 266 faillites en 1935 pour 1 085 en 1941; celui des liquidations judiciaires passe de 2 222 en 1937 à 48 en 1943 ! Pour la même année, 1 478 épiceries nouvelles s'ajoutèrent à celles qui existaient déjà. Jamais le commerce n'avait été aussi florissant !» (Hallier 1972 : 66)

Même si l'on ne doit pas accorder trop d'importance aux chiffres donnés par cet auteur, notamment à cause du changement intervenu dans le mode de calcul statistique, cette citation atteste de l'image très négative du petit commerçant. Les spécialistes du domaine confirment cette vision :

«Après la Seconde Guerre mondiale, le commerce avait en France une image plus négative que jamais. Le commerçant déjà suspect, comme toujours, de ne pas être moral car gagnant de l'argent sans être productif, c'est-à-dire sans participer à la transformation des biens, se voyait, en plus, accusé d'avoir activement participé au marché noir. À la Libération, quand on voulait insulter un individu ou désigner un parvenu, on le traitait de "BOF" (beurre, œufs et fromage).» (Benoun et Héliès-Hassid 1995 : 14; voir aussi Dutourd 1952)

«... le commerce intérieur, considéré comme un simple rouage technique et juridique dépendant n'a guère suscité d'analyses ou de recherches fondamentales jusqu'au lendemain de la Seconde Guerre mondiale (1). On peut même affirmer que la théorie économique, [...] les sciences humaines et l'université en général cachaient mal une certaine condescendance pour l'univers de la boutique, au reste partagée par une opinion publique manifestement méfiante à l'égard des "intermédiaires", la période 1939-1945 n'ayant pas contribué à atténuer ce sentiment.» (Coquery 1977 : 1- 2)

À ce ressentiment politico-social, il faut ajouter dans la période de l'immédiat après-guerre les reproches adressés aux commerçants pour leur côté traditionaliste et leur conservatisme, symbolisé, en France, par le poujadisme. Selon Ducrocq *et al.*, «en 1953, Poujade et le poujadisme confirment que le commerce français n'est pas animé par un esprit de progrès et d'innovation» (1994 : 10).

Religion	États-Unis	France
Catholiques	60 000 000	45 000 000
Protestants	76 000 000	785 000
dont calvinistes	6 000 000	90 000
Musulmans	?	2 000 000
Juifs	?	750 000

Voilà pour l'image peu glorieuse du commerçant. En changeant de nom, le «distributeur» Leclerc indique son intention de s'en démarquer, tant sur le plan éthique que purement commercial. Cette intention explique, au moins partiellement, l'apparition du nouveau paradigme lexical de la distribution. Certes, à ce facteur subjectif s'ajoute un fondement objectif, celui du changement structurel du commerce, lié à la distribution et à la consommation de masse, qu'avait amorcé Boucicaud en France avec la formule des grands magasins (du Closel 1989). Toutefois, le changement des valeurs semble lui aussi indéniable. La théorie économique du 20^e siècle en atteste. Si, au siècle dernier, le Littré rangeait encore le commerce dans les secteurs productifs : «Dans l'économie politique, l'agriculture, la fabrication, le commerce sont les trois branches de la production générale» (1873 : 684), au 20^e siècle, les économistes insistent sur l'image suspecte du commerçant, immoral puisqu'il gagne de l'argent sans être productif. L'I.N.S.E.E., l'office national français de la statistique, s'en est même fait l'écho puisqu'il rangeait jusqu'à récemment le commerce, système intermédiaire entre la production et la consommation, parmi les secteurs improductifs.

1.2 Apparition du paradigme de la distribution

Le divorce est ainsi consommé. La «distribution» est présentée par ses agents comme la solution des problèmes attachés au petit commerce. Ainsi, au «commerce sclérosé, fautif de goulets d'étranglement, aux marges trop élevées», s'opposent «ces industriels des temps modernes maîtrisant totalement l'écoulement de grandes quantités de produits» (Benoun et Héliès-Hassid 1995 : 13). À la boutique du commerçant, «symboliquement représenté par un crémier faisant les comptes des clients sur un carnet à souches avec un crayon qu'il dégage du coin de son oreille», s'oppose «un univers hautement technologique où règnent la lecture optique, l'informatique interactive et la manutention automatisée» et «demain la vente électronique [...] et le prélèvement automatique sur le compte bancaire ...» (Benoun et Héliès-Hassid 1995 : 13-14). Dans l'imaginaire populaire, cet univers impressionnant devient moins malhonnête, moins mesquin. Les distributeurs misent sur la productivité pour s'anoblir.

L'évolution technologique suffit-elle à garantir les valeurs éthiques ? Ce n'est certes pas l'objet de cet article d'en discuter, mais on ne peut pas ignorer les problèmes que soulèvent certaines des pratiques souvent en usage dans le milieu de la distribution : relations conflictuelles avec les fabricants, paiements à 90 ou à 180 jours, antisindicalisme, pressions sur les commissions d'urbanisme commercial. Toutefois, si la réputation de la distribution souffre parfois de telles pratiques, l'image auprès du public est différente, nettoyée des préjugés qui collent encore à la peau du commerçant, comme nous le verrons par l'étude des corpus de la presse actuelle. Mais au préalable, faisons une incursion dans le monde des dictionnaires pour voir comment {commerce, etc.} et {distribution, etc.} sont traités à travers les âges et quelles connotations s'attachent à ces termes et à leurs dérivés.

2. ÉTUDE ÉTYMOLOGIQUE ET LEXICOLOGIQUE DES PARADIGMES COMMERCE ET DISTRIBUTION

2.1 Les sens de commerce sont ambigus

Le mot commerce, qui vient du latin *commercium*, construit sur *cum*, avec et *merx*, marchandises (Dauzat *et al.* 1964; Rey *et al.* 1992), désigne «l'échange entre les

hommes des divers produits de la nature ou de l'industrie» (Littré 1873). Toutefois, il ne s'agit là que d'un sens parmi les cinq attestés dans le *Dictionnaire universel* de Furetière (1690). En effet, à ce premier sens neutre, dénotant une activité économique, s'ajoutent dès le 16^e siècle ceux de «relation réciproque» et de «trafic de choses morales». Signalons d'ailleurs que le *Dictionnaire de la langue française du seizième siècle* de Huguet (1925) ne donne «commerce» qu'au féminin, et uniquement dans le sens second de trafic des choses morales, avec une citation de Montaigne : «Nous qui n'avons aucune commerce avec eux. MONTAIGNE, II, 12, var. de 1588 (II, 166, et V, 122)» (Huguet, tome 2, 1925 : 367).

Le Littré (1873) relève les connotations négatives attachées au commerce, dans le sens de trafic des choses morales :

«Que vois-je autour de moi que des amis vendus,
Qui, choisis par Néron pour ce commerce infâme...»
(Racine, *Britannicus*, I,4).

Tous les dictionnaires consultés sont unanimes lorsqu'ils consignent les expressions «faire un mauvais / un méchant / un vilain commerce» pour dire «se mêler de quelque vilaine affaire», à côté des expressions neutres ou même positives telles que «le commerce des hommes» :

«Dans le brillant commerce il se mêle sans cesse
Et ne cite jamais que duc, prince ou princesse»
(Molière, *Le Misanthrope*, V,1.)

ou bien le «commerce des lettres» et le «commerce des muses» pour désigner les occupations littéraires. Il y a donc ambiguïté sur le sens de commerce, mot polysémique, comme le montre, entre autres, l'expression commerce sexuel, qui n'est pas l'équivalent de l'anglais *sex shop* et qui ne réfère pas aux pratiques que l'on sait, impliquant rémunération.

2.2 Le sens économique de commerce, une évolution

Même dans le sens économique, neutre, du mot commerce, les dictionnaires consultés reflètent une évolution de l'acception qui va faire qu'au 20^e siècle, le terme sera de plus en plus associé au petit commerce, sauf quand il garde son sens de discipline, comme dans «école de commerce», quand il garde son sens macro-économique, comme dans «ministère du Commerce», et dans les expressions «avoir la bosse / le sens du commerce».

Nous illustrons cette évolution par la recherche des phrasèmes haut commerce, grand commerce et petit commerce. Nous ne trouvons le phrasème haut commerce (construit sur le même modèle que haute finance) que dans deux dictionnaires : le *Dictionnaire général de la langue française du commencement du XVII^e siècle, jusqu'à nos jours* (Hatzfeld et Darmester 1964) et le *Grand Robert de la langue française* (1985). Ce dernier le consigne pour désigner l'ensemble des grands commerçants en spécifiant que c'est un usage vieilli. De fait, l'expression disparaît au 20^e siècle. Inversement, l'expression petit commerce, absente des dictionnaires anciens, est consignée dans tous les

ouvrages contemporains. Le phrasème grand commerce, quant à lui, se retrouve dans les dictionnaires des 18^e et 19^e siècles, comme dans l'exemple «le grand commerce de Mofcovie eft de fourrure» (*Nouveau Dictionnaire de l'Académie française*, éd. de 1718). Dans les ouvrages ultérieurs il perdure dans le sens de commerce international. Ainsi, il figure en bonne compagnie dans le *Trésor de la langue française*, comme le montre cette citation d'Aragon : «Toutes les élégantes sont là [...] ! La haute banque et le grand commerce, l'industrie en jaquette et en tube» (Aragon, *Les beaux quartiers*, 1936 : 491, cité dans le TLF, vol. 5, 1977 : 1119).

L'évolution lexicologique que nous avons relevée a sa source dans le glissement du sens du mot commerce, glissement qui s'opère à deux niveaux. Autrefois, «commerce» désignait exclusivement le commerce entre nations, le commerce international, comme dans les expressions commerce de l'Orient ou commerce des fourrures. Cette acception se reflète dans la syntaxe du verbe commercer, attesté dès 1470 : «on commerce la soie quelque part», impliquant donc voyage, commerce extérieur. La réalité intérieure (nationale) du «petit commerçant», celle de la boutique ou de l'échoppe, était désignée par les termes marchands et artisans. Ce n'est qu'au 20^e siècle qu'apparaissent, au moins dans les dictionnaires, la distinction entre commerce intérieur et commerce extérieur et encore plus récemment la notion d'import-export. L'apparition de ces syntagmes libère l'usage de commerce en le restreignant pour désigner principalement le petit commerce.

De plus, en ce qui concerne le commerce intérieur, à cause de l'évolution même du monde commercial, apparaît une autre distinction entre petit commerce et distribution (ou grande distribution; voir *supra*). L'étude diachronique des termes montre donc une restriction des sens de commerce, ainsi que l'apparition de connotations négatives attachées au concept de commerce. Par exemple, l'adjectif commercial s'entache d'un sens péjoratif — «conçu, exécuté dans une intention lucrative, [...] et pour plaire au plus grand public possible» —, consigné en troisième position (sur cinq acceptions) dans le *Petit Larousse* avec les exemples jazz commercial et peinture commerciale, pour dire «n'étant pas de qualité».

2.3 Le sens de distribution

Alors que pour «commerce», le premier sens est économique, pour «distribution», c'est le contraire. Le terme est général au départ, puis il prend des sens spécifiques (en droit, en imprimerie) et n'apparaît avec le sens moderne «opérations et circuits grâce auxquels les biens de consommation sont acheminés vers les points de vente» que très récemment. Au premier sens consigné à l'entrée distribution, le *Grand Robert de la langue française* renvoie à dons, largesse, libéralité (1985 : 594), au moyen de deux exemples : «Dieu, qui est roy et distributeur des royaumes» et «Dieu distribue la lumière». Ainsi, le sens commercial de distribution n'est consigné ni dans le *Petit Robert* de 1990 ni dans le *Dictionnaire historique de la langue française* (1992). Par contre, il figure dans le *Grand Robert de la langue française* (1985) avec un développement encyclopédique. Le *Trésor de la langue française* (vol. 7, 1979) se montre plus réticent à consigner l'usage moderne du mot distribution puisqu'il n'en donne pas de définition et l'insère au moyen de citations (en précisant distribution commerciale).

Si le sens commercial du terme distribution n'apparaît pas dans le *Petit Larousse illustré* de 1992, il figure par contre dans celui de 1998, avec une définition appropriée et

la collocation grande distribution. Par contre, le dérivé distributeur n'y figure pas encore dans le sens de société de distribution (enseigne), pourtant très employé dans le milieu professionnel. Le *Petit Larousse*, en avance sur les autres dictionnaires, consigne toutefois le sens commercial un peu plus général de distributeur : «personne ou firme qui assure la distribution d'un produit, service, film». On peut donc dater avec une certaine précision l'apparition du sens commercial du triplet {distribution, distributeur, distribuer} et relever le retard des dictionnaires de langue générale sur la réalité, mais aussi sur les encyclopédies, puisqu'une ancienne édition de l'*Encyclopaedia Universalis* comporte déjà un développement de cinq pages sous l'entrée distribution (1985 : 287-291).

Quant aux connotations de distribution, dans ses sens généraux, telles qu'elles apparaissent dans les citations ou les exemples donnés dans les dictionnaires, elles sont toutes positives ou neutres : distribution des revenus, des richesses, du sang dans les artères, des prix, des lettres (le distributeur de lettres); distribuer la communion. On trouve aussi, par plaisanterie, distribuer la raclée (Camus, *La chute* 1956, cité dans le TLF, vol. 7, 1979 : 333) ou «la vieille servante, [...] grande distributrice de taloches et de fessées» (Zola, *Vérité*, 1902, cité dans le TLF, vol. 7, 1979 : 333)

3. ÉTUDE LEXICOMÉTRIQUE DES PARADIGMES COMMERCE ET DISTRIBUTION DANS LA PRESSE CANADIENNE ET FRANÇAISE

L'idée de base consistait à vérifier dans des sous-ensembles de textes extraits de la presse canadienne et française actuelle si les connotations attachées aux formes du paradigme commerce étaient plus nombreuses que celles attachées au paradigme distribution. Pour ce faire, nous avons constitué les corpus «commerc» et «distrib» en utilisant des textes saisis dans la banque de données TEXTUM de l'Université de Montréal. Le contenu de cette banque a déjà été décrit ailleurs (Clas 1996 : 135-137). Nos deux corpus ont été divisés en plusieurs sous-corpus, que nous continuerons toutefois d'appeler corpus dans la suite de cet article (pour des raisons de légèreté évidentes). «Commerc» est donc composé de «commercmo» (*Le Monde*), «commercfof» (*Ouest-France*) et «commercanc» (*Presse canadienne*). De même, «distrib» se divise en trois : «distribmo», «distribfof» et «distribanc». Ensuite, nous avons segmenté en unités minimales chacun des six nouveaux corpus avec le logiciel Lexico1 et nous avons inséré une clé de partition dans chaque corpus. Nous y avons alors recherché les concordances des formes dérivées des séries {commerce...} et {distribution...}, dans lesquelles nous avons sélectionné les contextes qui révélaient les connotations négatives ou positives attachées aux formes des deux paradigmes (recherche de cooccurrences lexicales, au sens lexicométrique; voir Lafon 1981).

Une étude préalable de la ventilation comparée des différentes formes du paradigme commerce dans nos trois corpus du *Monde*, de *Ouest-France* et de la Presse canadienne indiquait que le lemme² commerce représente entre 41 et 46 % des occurrences du paradigme commerce dans les trois corpus, alors que le lemme commerçant n'en représente que 8 à 19 %. Notons aussi l'emploi rare du verbe commercer (0,1 à 0,5 %), pris en

² Le lemme commerce regroupe tout simplement les formes commerce et commerces. De même, le lemme commercer regroupe les formes conjuguées de ce verbe. Cela dit, nous tenons à préciser que nous ne travaillons pas avec un corpus lemmatisé. Nous avons seulement procédé à un «calcul lemmatisé» ponctuel manuel.

général dans le sens de faire du commerce avec l'étranger. Le même phénomène se reproduit avec le verbe distribuer, beaucoup plus rare que distribution. Le tableau qui suit résume les cooccurrences (commerce; x) dans la concordance de commerce établie à partir du corpus commercmo (*Le Monde*) :

1.	accord général sur les tarifs douaniers et le commerce	165
2.	chambre de commerce (et d'industrie)	231
3.	commerce de détail	23
4.	commerce de gros	3
5.	commerce extérieur	474
6.	commerce intérieur	5
7.	commerce international	230
8.	commerce mondial	104
9.	département (américain) du commerce	74
10.	écoles (supérieures) de gestion et de commerce/ grandes écoles de commerce (et d'ingénieurs)	94
11.	fonds de commerce	65
12.	grand commerce	12
13.	ministère de l'industrie et du commerce extérieur/du commerce et de l'artisanat/du commerce international et de l'industrie	46
14.	ministre (délégué au) du commerce international et de l'industrie /commerce (extérieur)/artisanat	217
15.	tribunal de commerce	175

Tableau 1 - Cooccurrences de commerce dans *Le Monde* (F = 2748)

Ce tableau de cooccurrences appelle trois remarques. La première concerne la primauté des occurrences de commerce extérieur au détriment de commerce intérieur. Ainsi, on observe cinq occurrences de l'expression commerce intérieur, contre environ 800 pour l'ensemble commerce extérieur, commerce international et commerce mondial. Dans la presse canadienne, le tableau des cooccurrences est très semblable, avec pour seule différence notable, une fréquence très inférieure pour l'expression école de commerce. La deuxième remarque porte sur les sens de commerce dans ces contextes, sens qui sont dans la plupart des cas neutres, macroéconomiques, notamment lorsque le commerce désigne le secteur d'activité ou la discipline enseignée. Enfin, notons que les connotations négatives sont absentes de ce relevé de cooccurrences.

Ainsi, la concordance du terme commerce ne nous permet pas de déceler les connotations négatives que nous supposons, au début de cette étude, attachées au petit commerce. Conscients du caractère exploratoire de notre recherche, nous devons ultérieurement approfondir les raisons de cette absence, notamment en utilisant d'autres

corpus. En effet, humbles et prudents devant les chiffres, nous nous gardons de toute extrapolation hâtive et risquée (à ce sujet, voir Tournier 1980).

L'étape suivante consistait à rechercher les cooccurrences de «commerçant». C'est là que nous avons pu valider en partie notre hypothèse sur les connotations négatives qui expliqueraient, entre autres facteurs, le changement de paradigme de commerce à distribution, et de commerçant à distributeur. Les graphes suivants représentent les réseaux associatifs des cooccurrences de «commerçant» dans les corpus commercmo (France) et commercan (Canada). Nous avons regroupé les contextes canadiens (405) et français (125) que comporte notre concordance, et avons étiqueté (empiriquement) les sept catégories qui se dessinaient : (1) classe sociale; (2) qualificatifs (positifs ou négatifs); (3) actions (positives ou négatives); (4) problèmes (créés ou subis par les commerçants); (5) perceptions de la société (clichés); (6) mouvements-regroupements; (7) origine ethnique. Ces catégories contribuent toutes à révéler les connotations associées à «commerçant». De plus, celle de la classe sociale permet de situer le commerçant par rapport à d'autres groupes sociaux.

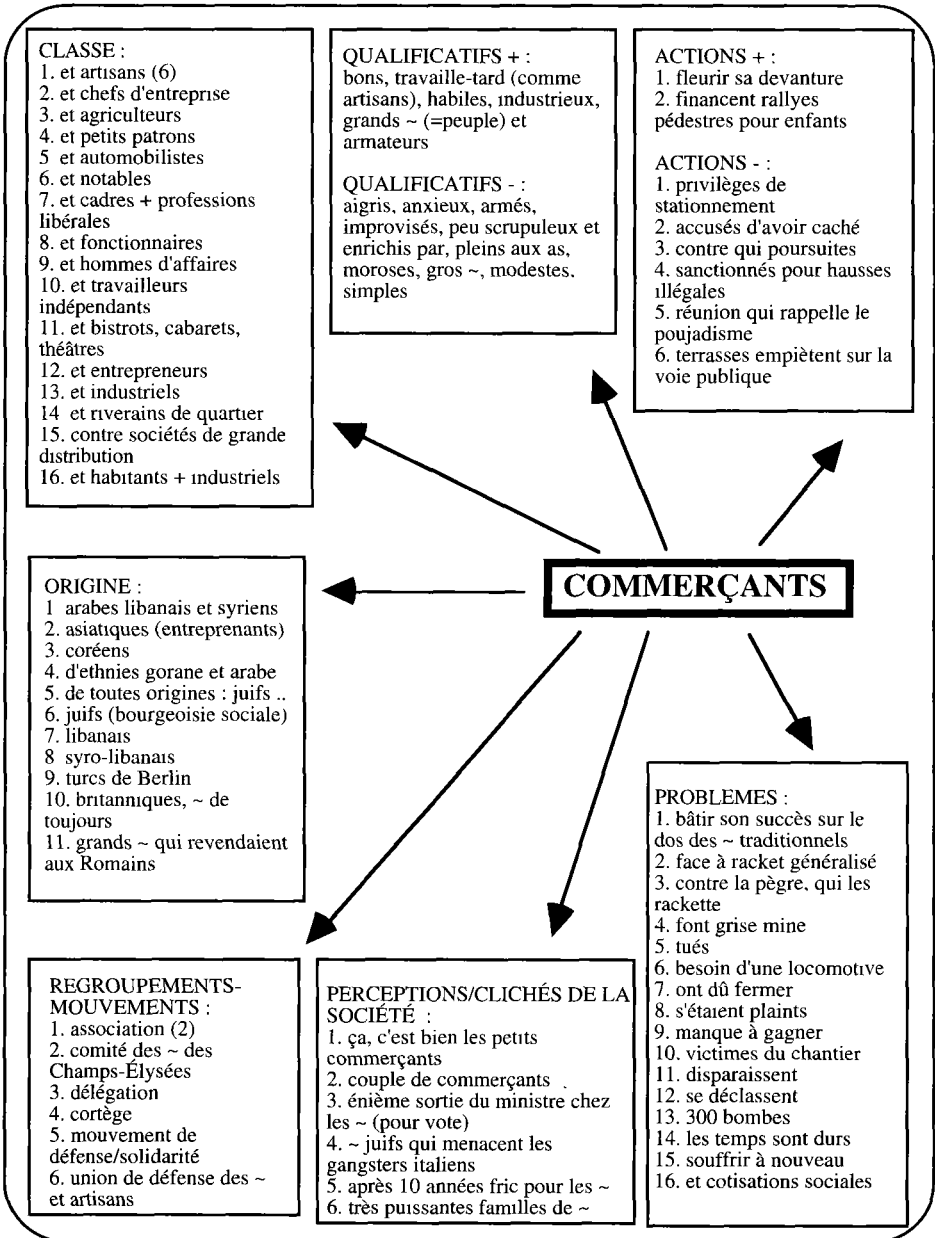


Fig. 1 - Réseau associatif des cooccurrences de commerçants dans le corpus du Monde (commercmo)

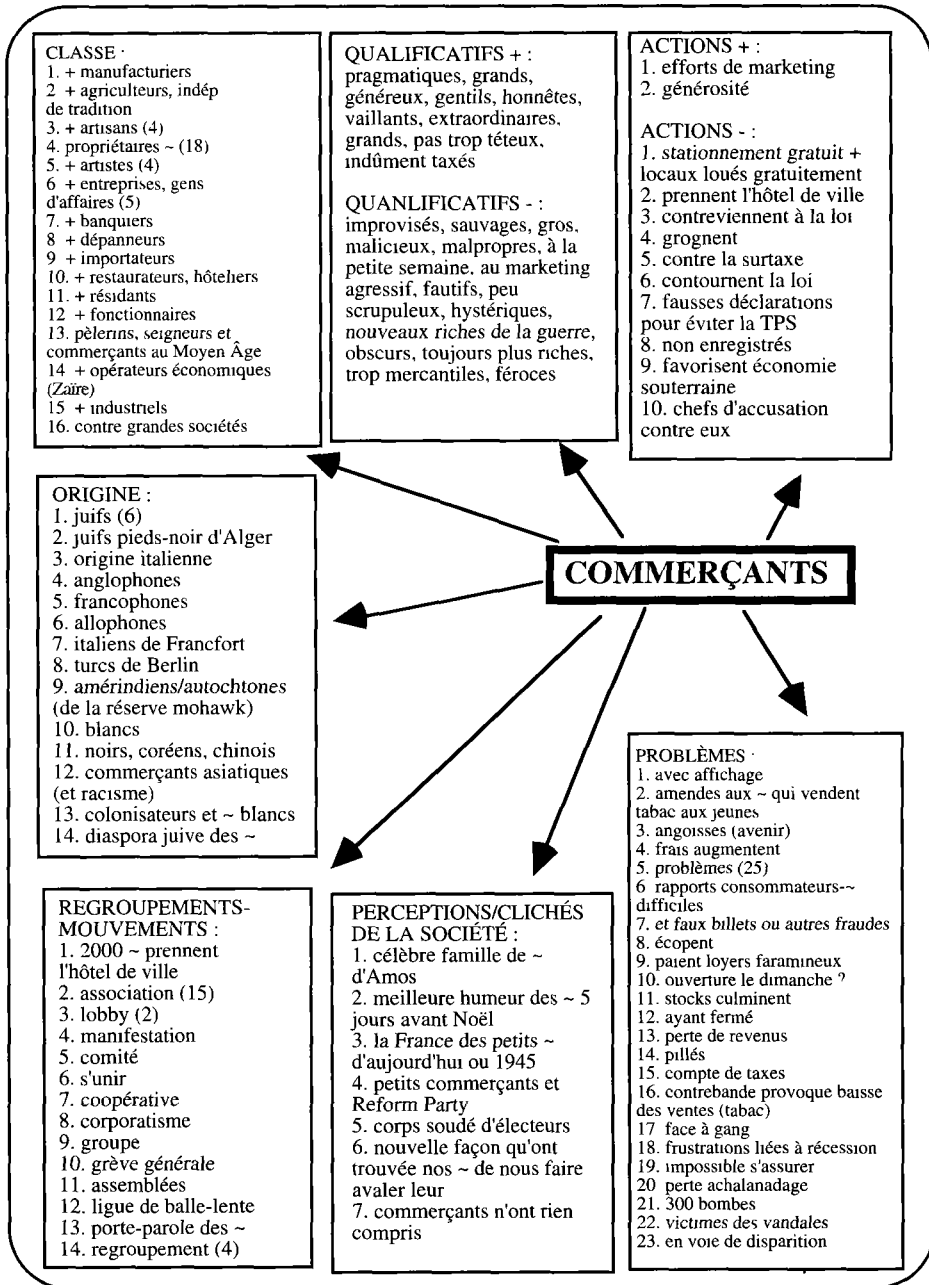


Fig. 2 - Réseau associatif des cooccurrences de commerçants dans le corpus de la presse canadienne française (commercan)

Les réseaux de cooccurrences ou réseaux associatifs permettent d'opérer la jonction entre statistique, terminologie et analyse du discours. Ces outils, très porteurs, sont de plus en plus utilisés (voir Debarbieux et Ravix 1993 : 554-555; Bourion 1995). Les cooccurrences, étudiées catégorie par catégorie, révèlent les associations faites avec «commerçant» dans la presse. Nous pensons que ces associations ne sont pas le fruit du hasard — un hasard théorique qui ferait que les unités de la langue auraient une distribution équiprobable —, mais qu'elles contribuent à forger (ou révéler) l'image du commerçant. Ainsi, pour ne prendre que quelques exemples, nous repérons, dans la catégorie classe, la cooccurrence (positive) de commerçants et artisans. Mais dans les qualificatifs, nous remarquons que les éléments négatifs sont plus nombreux que les positifs. Dans les deux catégories mouvements-regroupements et perceptions de la société, nous remarquons que le côté associatif des commerçants est perçu de façon négative, comme le montrent la référence explicite au poujadisme dans le corpus français et les références au corporatisme dans la presse canadienne française.

La catégorie origine est certainement très révélatrice. Les qualificatifs de nationalité (commerçants juifs, asiatiques, turcs) traduisent un phénomène connu, peut-être basé sur un fondement objectif mais aussi sur des préjugés ou une image sociale stéréotypée : la propension des émigrés à ouvrir un commerce lorsqu'ils arrivent à l'étranger (commerçants turcs à Berlin, italiens à Francfort, portugais de la rue Prince-Arthur de Montréal, etc.). Mais il est intéressant de relever que les nationalités «non marginales» (c'est-à-dire majoritaires) ne sont jamais notées. Ainsi, nous ne trouvons pas commerçants «français», ni «américains», ni «allemands». Dans le corpus canadien, nous retrouvons les marques de deux réalités propres au Québec : les commerçants francophones par opposition aux anglophones ou aux allophones, à cause des problèmes liés aux règlements sur l'affichage, et les cooccurrences de commerçants avec autochtones et amérindiens (dans les réserves) associées aux avantages fiscaux dont ils semblent jouir, selon la presse et l'opinion publique. Quant à l'étude parallèle des cooccurrences de distributeur, elle n'a révélé aucune connotation négative.

CONCLUSION

Nous avons présenté dans cet article certains des résultats tirés d'une étude lexicométrique sur les triplets {commerce, commerçant, commercer} et {distribution, distributeur, distribuer}, en mettant l'accent sur le contraste France-Québec et sur les couples commerce-commerçant et distribution-distributeur. En raison des contraintes matérielles imposées dans cet article, les fréquences inférieures des verbes commercer et distribuer justifiaient ce choix. Cette étude exploratoire devait apporter des éléments de réponse à la question que nous posions sur les connotations négatives comme facteur de changement du paradigme lexical commerce-commerçant à celui de distribution-distributeur.

Nous sommes bien conscients des limites de telles études pour vérifier des phénomènes de langue complexes : un corpus ne révèle que ce qu'il contient. Les articles de journaux sont révélateurs de la presse écrite, d'une volonté d'information sur des sujets d'intérêt général, voire national. Ils sont inscrits dans le temps présent et peuvent faire l'objet d'une auto-censure. De plus, nous n'avons pu incorporer dans notre corpus les articles de journaux datant des périodes «sensibles» telles que l'immédiat après-guerre.

Néanmoins, à l'issue de cette étude et malgré les limites que nous venons d'énoncer, quelques remarques s'imposent. D'une part, les réseaux associatifs du terme commerçant font nettement ressortir les images négatives attachées à ce terme. Ces résultats vont dans le sens de l'hypothèse que nous avons émise comme explication du changement d'appellation chez les grands commerçants qui, pour se distinguer des petits commerçants, adoptent le terme distributeur. D'autre part, nous sommes surpris par le peu de différence, sur le plan quantitatif, entre les listes de connotations négatives au Québec et en France. En supposant que les commerçants installés au Québec avaient un passé moins lourd que leurs homologues français (collaboration), on aurait pu s'attendre à ce que les connotations négatives qui pèsent sur la classe des commerçants fussent moins nombreuses en Amérique du Nord. Mais l'eau est passée sous les ponts depuis la Seconde Guerre mondiale. Notre étude des connotations montre que les préjugés ressortent de toutes les façons sans qu'il y ait d'événements dramatiques les nourrissant. Nous avons également montré que la Seconde Guerre mondiale n'a été qu'un des facteurs, parmi d'autres, qui ont contribué à ternir la réputation des commerçants.

Sur un plan plus général, nous pensons avoir illustré dans cette étude des paradigmes commerce et commerçant le rôle, souvent inconscient, des connotations comme facteur déterminant d'évolution terminologique. Les mots ont une mémoire !

RÉFÉRENCES

- Académie française (1994) : *Nouveau dictionnaire de l'Académie française*, édition de 1718, Genève. Slatkine Reprints.
- AL-HARÎRÎ, A. (1992) : *Le Livre des malias, séances d'un vagabond de génie*. Phébus, Paris.
- ATTALI, J. (1991) : *1492*, Paris, Fayard.
- BALZAC, H. de (1963) : *La Maison du Chat-qui-pelote; Le bal de Sceaux; La Vendetta*, collection «Classiques Garnier», Paris, Garnier.
- BALZAC, H. de (1975) : *César Birotteau*, édition de 1990, collection «Lire et voir les classiques», Paris, Pocket.
- BENOUN, M. et M.-L. HÉLIÈS-HASSID (1995) : *Distribution, Acteurs et stratégies*, 2^e éd., collection «Gestion», Paris, Économica.
- BOURION, É. (1995) : «Le réseau associatif de la peur». F. Rastier (dir), *L'analyse thématique des données textuelles. L'exemple des sentiments*, collection «Études de sémantique lexicale», Paris, Didier Érudition.
- BRAUDEL, F. (dir) (1985) : *La Méditerranée : l'espace et l'histoire*, collection «Champs», Paris, Flammarion.
- CANTIN, A. (1996) : «Négociation : la méthode secrète du roi Carrefour», *L'Essentiel du Management*, n° 19, p. 82-90.
- CHARMES, C. (1989) : *Cours de politique générale*, fascicule 2 (Politique générale et culture), Angers (France), ESSCA.
- CHUNG, C.H., SHEPARD, J.M et M.J. DOLLINGER (1989) : «Max Weber Revisited : Some Lessons from East Asian Capitalistic Development», *Asia Pacific Journal of Management*, vol. 6, n° 2, pp. 307-321.

- CLAS, A. (1996) : «Dictionnaires généraux bilingues : Le *Dictionnaire bilingue canadien*», *Cahiers de lexicologie*, n° 69, pp. 127-144.
- CLOSEL, J. du (1989) : *Les grands magasins français. Cent ans après*, collection «Connaissance du commerce et de la consommation», Paris, ICC et Chotard et Associés Éditeurs.
- COQUERY, M. (1977) : *Mutations et structures du commerce de détail en France : étude géographique*, Cergy Village (France), Éditions du Signe.
- DANCETTE, J. et C. RÉTHORÉ (1997) : *Dictionnaire bilingue de la distribution/Bilingual Dictionary of Retailing*, version préliminaire, Université de Montréal, Département de linguistique et de traduction.
- DANCETTE, J. et C. RÉTHORÉ (1997) : «Le Dictionnaire bilingue de la distribution : entre dictionnaire de langue et encyclopédie. *Meta*, vol. 42, n° 2. Communication présentée aux IV^{es} Journées scientifiques du réseau «Lexicologie, terminologie, traduction» de l'AUPELF-UREF (Lyon, France, 28-30 septembre 1995).
- DAUZAT, A., DUBOIS, J. et H. MITTERAND (1972) : *Nouveau dictionnaire étymologique et historique*, 3^e éd. rev. et corr., Paris, Larousse.
- DAVID, M. (1995) : «Il était une fois le nouveau commerce», *LSA*, n° 1000.
- DEBARBIEUX, B., et B. RAVIX (1993) : «Le territoire en discours : Saint-Gervais-les-Bains», *Actes des secondes journées internationales d'analyse statistique de données textuelles*, Paris, S.J. Anastex Éditeur, pp. 551-562.
- DUCROCQ, C., JAMIN, N. et S. LAGRANGE (1994) : *La distribution*, collection «Vuibert Entreprise», Paris, Vuibert.
- DUTOURD, J. (1952) : *Au bon beurre : ou, dix ans de la vie d'un crémier*, Paris, Folio.
- Encyclopaedia Universalis* (1995-1996) : Paris, Encyclopaedia Universalis.
- FAUVEAU, J.-C. (1991) : *Le monde de la distribution*, Noiziel (France), Les Presses du Management.
- FAWCETT, H. (1989) : *A Text Searching System. PAT 3.3. User's Guide*, Waterloo, UW Centre for the New Oxford English Dictionary.
- FURETIÈRE, A. (1690) : *Dictionnaire universel contenant généralement tous les mots françois tant vieux que modernes, et les termes de toutes les sciences et des arts...*, La Haye, Leers.
- HALLIER, J.-E. (1972) : *La cause des peuples*, Paris, Seuil.
- HARDOUIN, P. et C. DUCROCQ (1993) : *Le commerce indépendant organisé. Rapport du groupe de travail de l'ICC*, collection «Connaissance du commerce et de la consommation», Paris, ICC et Éditions Liaisons.
- HATZFELD, A., et A. DARMESTER (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*, Paris, Delagrave.
- HUGUET, E. (1925) : *Dictionnaire de la langue française du seizième siècle*, Paris, Champion.

- IMBS, P. (dir) (1971-1994) : *Trésor de la langue française : dictionnaire de la langue du XIXe et du XXe siècle (1789-1960)*, Paris, Éditions du CNRS
- LAFON, P. (1980) : «Sur la variabilité de la fréquence des formes dans un corpus», *Mots*, n° 1, pp. 127-165.
- LAFON, P. (1981) : «Analyse lexicométrique et recherche des cooccurrences. Des rapports syntagmatiques aux cooccurrences», *Mots*, n° 3, pp. 95-148.
- LAFON, P. (1984) : *Dépouillements et statistiques en lexicométrie*. Genève-Paris, Slatkine-Champion.
- LEBART, L. et A. SALEM (1994) : *Statistique textuelle*, Paris, Dunod.
- Les Mille et Une Nuits. Histoire de Sindbad le Marin*, trad. d'Antoine Galland (1996) : collection «Librio», Paris, Flammarion.
- LITTRÉ, É. (1873) : *Dictionnaire de langue française*. Paris, Hachette.
- Petit Larousse illustré* (1992) : Paris, Larousse.
- Petit Larousse illustré* (1998) : Paris, Larousse.
- REY, A. et al. (1992) : *Dictionnaire historique de la langue française*, Paris, Dictionnaires Le Robert.
- ROBERT, P. (1985) : *Grand Robert de la langue française*, 2^e éd., Paris, Le Robert.
- ROBERT, P. (1990) : *Dictionnaire alphabétique et analogique de la langue française : le Petit Robert I*. Paris/Montréal, Le Robert/Robert-Canada.
- ROBERT, P. (1998) : *Dictionnaire alphabétique et analogique de la langue française : le Petit Robert I*, Paris/Montréal, Le Robert/Robert-Canada.
- SALEM, A. (1990) : *LexicoI*, version 0.7, Saint-Cloud (France), Lexicométrie et textes Politiques, INaLF-CNRS/ENS de Fontenay-St.Cloud.
- SALEM, A. (1992) : *LexicoI-Version 0.7. Notice d'utilisation*. Saint-Cloud (France). Lexicométrie et textes Politiques, INaLF-CNRS/ENS de Fontenay-St.Cloud.
- SÉDILLOT, R. (1964) : *Histoire des marchands et des marchés*. Paris, Fayard.
- THIL, E. (1966) : *Les inventeurs du commerce moderne*, Paris, Arthaud.
- TOURNIER, M. (1980) : «D'où viennent les fréquences de vocabulaire ?», *Mots*, n° 1, pp. 189-212.
- WEBER, M. (1964) : *L'éthique protestante et l'esprit du capitalisme*, traduit de l'allemand par J. Chavy, Paris, Plon.
- YDEWALLE, C. d' (1965) : *Au Bon Marché, de la boutique au grand magasin*, Paris, Plon.
- ZOLA, É. (1990) : *Au Bonheur des Dames*, édition de 1883, collection «Lire et voir les classiques», Paris, Pocket.

NÉOLOGISMES ET TERMES D'EMPRUNT DANS LE FRANÇAIS DES AFFAIRES

Une étude de la terminologie en usage dans les offres d'emploi de la presse française

Jean SOUBRIER

INSA de Lyon, France

Depuis 1972, date à laquelle fut créée au ministère de l'Économie et des Finances la toute première commission de terminologie, pas moins de six arrêtés¹ relatifs à la terminologie économique et financière ont été adoptés par les gouvernements successifs. Ces arrêtés, qui rendent obligatoires ou fortement conseillés plus de 250 termes dans les documents officiels ont pour but avoué de faire disparaître de la langue des affaires bon nombre de termes empruntés à l'anglo-américain. Largement relayés par la presse et par divers organismes publics ou privés, les travaux des terminologues ont bien sûr contribué à l'enrichissement du français avec un certain succès. Dans quelle entreprise n'utilise-t-on pas quotidiennement aujourd'hui des mots comme *crédit-bail*, *logiciel*, *départ-usine* et *savoir-faire* au lieu de *leasing*, *software*, *ex-works* et *know-how*, qui étaient les seuls termes en usage voici vingt ans ? Toutefois ces réussites terminologiques ne doivent pas occulter la réalité des faits. Les termes de la langue des affaires empruntés à l'anglo-américain demeurent toujours très prisés dans les milieux professionnels. Il n'est pour s'en convaincre que de lire les titres des ouvrages exposés au rayon «entreprise», d'une librairie ou, plus simplement encore, de parcourir les pages économiques et commerciales de la presse française. Pourtant, en l'absence de données chiffrées, notre appréciation du phénomène ne peut être que subjective et c'est pourquoi nous avons décidé d'entreprendre une étude sur la terminologie en usage dans les offres d'emploi publiées dans la presse pour cadres. Ce choix s'est imposé à nous pour différentes raisons alliant des exigences d'ordre pratique à des considérations plus théoriques.

Rédigées à l'intention du candidat idéal, les offres d'emploi concentrent en quelques mots l'image que l'entreprise veut donner d'elle-même. Elles constituent à ce titre un terrain d'observation privilégié. Par ailleurs il faut souligner que le support que nous avons retenu (presse généraliste) publie des annonces couvrant de très nombreux secteurs du monde professionnel et cela nous a permis d'avoir une vision assez large du phénomène. Enfin, et ce n'est pas la moindre de nos préoccupations, les offres d'emploi, publiées selon un rythme régulier et facilement repérables au coeur d'un magazine simplifient grandement la constitution d'un corpus.

¹ Arrêtés du : 29 novembre 1973, 18 février 1987, 6 janvier 1989, 11 janvier 1990, 30 septembre 1991, 11 février 1993.

Cette étude, qui ne porte que sur un domaine très circonscrit de la langue des affaires, s'est fixé un triple objectif :

- Repérer et quantifier les emprunts à l'anglo-américain dans les offres d'emploi.
- Vérifier l'hypothèse selon laquelle les néologismes « officiels » ne connaissent qu'une très faible diffusion auprès des utilisateurs professionnels.
- Analyser, dans une perspective linguistique mais également culturelle, les principaux facteurs qui freinent la diffusion des néologismes dans le monde des affaires et qui contribuent de ce fait à l'utilisation plus ou moins généralisée d'une terminologie d'emprunt.

PRÉSENTATION DE L'ÉTUDE

Afin de constituer notre corpus, nous avons choisi le supplément « Réussir » publié chaque semaine par le magazine *L'Express*. Ce supplément détachable, qui contient en moyenne plus d'une centaine d'offres d'emploi, présente en outre l'avantage d'appartenir à un très vaste service d'offres d'emploi accessible sur Internet auquel nous avons eu recours ultérieurement. Nous avons procédé au dépouillement de 12 numéros successifs (n° 2389 au n° 2400 entre le 17 avril et le 3 juillet 1997) et étudié 1473 offres d'emploi rédigées en français (les quelques offres d'emploi rédigées en anglais n'ont bien sûr pas été prises en compte).

MÉTHODOLOGIE

La première question qui s'est posée à nous fut celle du repérage des emprunts à l'anglais dans notre corpus. Sans rentrer dans le détail de la problématique du repérage qui reste fondamentalement liée à toute activité terminologique² nous résumerons ici les choix que nous avons faits :

« Quand on parle d'emprunt linguistique, c'est d'abord aux mots que l'on pense »³ et bien naturellement nous avons relevé tous les termes dont la graphie trahissait à l'évidence l'importation lexicématique, c'est-à-dire « l'emprunt de signe complet à la fois comme expression et comme contenu »⁴ par exemple : *leader, start up, sourcing, challenge*, etc.

Nous avons aussi relevé quelques graphies hybrides manifestement situées entre anglais et français comme *merchandising, mixt, télémarketing, entrepreneurial* ou encore nées d'une inversion de deux termes comme *mix marketing*.

« C'est avoir une idée bien superficielle de l'anglicisme que de ne le considérer que comme un mot emprunté. D'une part nous n'empruntons pas que des mots, et d'autre part il ne s'agit pas toujours du signe complet mais de son expression ou de son contenu

² CHUKWU, Uzoma (1993): *Le repérage des termes dans un corpus bilingue anglais/français*. Thèse de Doctorat dirigée par Ph. Thoiron, Faculté des langues, Université Lumière Lyon 2, 545p.

³ DEROY, Louis (1956): *L'emprunt linguistique*, Bruxelles, Sté. d'édition Les belles lettres, p.67.

⁴ HÖFLER, Manfred (1982) : *Dictionnaire des anglicismes*, Paris, Larousse, p.VI.

seulement»⁵. Cette remarque de Josette Rey-Debove nous a incité à relever, en marge de notre liste principale, une autre série d'emprunts traditionnellement décrits comme des anglicismes clandestins ou anglicismes sémantiques. Nous avons ainsi réuni des mots (cf. Annexe B) ayant déjà fait l'objet d'un débat ancien (ex : *contrôle*, *opportunité*, *approche*, *attractif*) et répertoriés dans de nombreux ouvrages⁶, mais aussi des mots d'introduction beaucoup plus récente qui témoignent de l'évolution de notre langue (*developpeur* = responsable de l'accroissement du volume d'affaires, *junior* = débutant, *senior* = confirmé, *être en charge* = être responsable de, etc.). Toutefois, nous avons pris le parti de ne pas comptabiliser ces emprunts sémantiques dans notre liste principale dans la mesure où il n'y a pas véritablement d'importation lexicématique mais seulement importation de structure sémantique.

Nous n'avons pas non plus retenu dans notre liste principale les emprunts à l'anglais qui ont été «officiellement» francisés après consultation des diverses commissions de terminologie. Ces emprunts officiels sont des termes dont la forme initiale est parfaitement compatible avec notre système morpho-phonologique et qui, de ce fait, ne posent pas de réel problème d'assimilation (ex. : *management*, *maintenance* etc.).

Nous n'avons pas davantage retenu les dérivations d'emprunts totalement francisées dans leur forme (ex. : «vous *managerez* un centre de profit...» ou «vos qualités *managériales*...») même si celles-ci sont parfaitement contestables comme, par exemple, le terme *profitabilité*.

RÉSULTATS

Au terme de notre phase de repérage nous avons pu établir une liste de 142 emprunts lexicaux à l'anglais (mots simples ou mots composés) relevés dans 1473 offres d'emploi. Cette liste que nous présentons sous forme de tableau en Annexe A, appelle un certain nombre de commentaires.

L'élément le plus important de cette liste figure dans les chiffres placés en regard de chaque entrée. Ces chiffres indiquent les occurrences enregistrées pour chacun des emprunts. Ces occurrences varient de 542 à 1. Mais le fait le plus surprenant est que deux mots, *marketing* et *leader*, totalisent à eux seuls près de 65 % des occurrences relevées.

Inversement, on constate que plus de la moitié des emprunts relevés (83 sur 142) n'ont qu'une seule occurrence dans notre corpus.

Par ailleurs nous avons cherché à quantifier la fréquence moyenne d'emploi des termes d'emprunts dans l'ensemble de notre corpus.

⁵ REY-DEBOVE, J. et G. GAGNON (1980) : *Le dictionnaire des anglicismes*, Paris, Le Robert, p. IX.

⁶ REY-DEBOVE, J. et G. GAGNON (1980) : *Le dictionnaire des anglicismes*, Paris, Le Robert. WALTER, Henriette et Gérard WALTER (1991) : *Dictionnaire des mots d'origine étrangère*, Paris, Larousse.

WALTER, Henriette (1997) : *L'aventure des mots venus d'ailleurs*, Paris, Robert Laffont.

N'ayant pas eu la possibilité matérielle de procéder à un traitement informatique complet (utilisation d'un scanner, transcription en fichier texte et comptage automatique des mots), nous avons dû procéder à une estimation moyenne à partir d'un comptage manuel portant sur 147 annonces sélectionnées de manière aléatoire, ce qui correspond environ à 10 % de l'ensemble des offres dépouillées. Ce comptage, qui n'a porté que sur le seul texte des annonces et pas sur les raisons sociales et les adresses des annonceurs, nous a permis d'évaluer la taille de notre corpus à 207 000 mots. Rapproché du nombre total d'occurrences (1619) cumulées (c-à-d. calculé en prenant en compte tous les mots à l'intérieur d'une entrée composée) nous obtenons un chiffre de 1 emprunt tous les 128 mots ce qui, traduit en pourcentage, nous donne une proportion moyenne d'emprunts égale à 0,78 % sur l'ensemble du corpus.

Ce chiffre, qu'il convient bien sûr de considérer comme une simple estimation, est néanmoins loin de correspondre au déferlement d'américanismes que l'on nous annonce régulièrement depuis le célèbre pamphlet d'Etienneble⁷. Il semble même à peine plus élevé que la proportion des anglicismes relevés il y a vingt ans dans le journal *Le Monde* et que l'on estimait à 0,6 % (soit un emprunt tous les 166 mots)⁸.

Si l'on considère de surcroît que la langue des affaires compte parmi les langues de spécialité dont la terminologie a été le plus influencée par les emprunts à l'américain nous pourrions même trouver dans ces chiffres un motif de satisfaction quant à la capacité du français à ne pas dépasser un certain «seuil de tolérance» dans le volume de ses importations.

Plus encore, le fait que seuls 11 termes de notre liste enregistrent des occurrences supérieures à 10 nous invite à relativiser davantage la portée du phénomène ainsi que le faisait déjà Jacques Cellard en écrivant : «Dans l'ensemble, il en est de l'investissement de notre langue par l'anglais ce qu'il en est, révérence parler, des figurants du Châtelet : en en faisant passer et repasser une douzaine sur scène, on donne à bon compte l'illusion d'une armée. Peu ou prou on en revient toujours aux mêmes...»⁹.

Étant ainsi parvenu à une estimation chiffrée de l'anglicisation de la langue des offres d'emploi, nous avons ensuite cherché à vérifier l'hypothèse selon laquelle les néologismes proposés par les commissions de terminologie n'étaient que très rarement employés par les professionnels du commerce et de l'industrie. Naturellement nous avons rencontré certains de ces néologismes au cours de la première partie de notre travail mais dans des proportions si réduites qu'il nous apparaissait alors dérisoire d'en faire état. Nous avons donc choisi d'adopter une démarche inverse et d'utiliser à des fins terminologiques les ressources d'Internet.

Ainsi nous avons eu recours au serveur «Cadres on line» (<http://www.cadronline.com>) qui regroupe les offres d'emploi publiées dans 17 titres de

⁷ ETIEMBLE, René (1964) : *Parlez-vous français ?*, Paris, Gallimard.

⁸ KLEIN, Virginia (1977) : *A study of Franglais in twelve issues of Le Monde*, (Mémoire de maîtrise dirigé par Guy J. Forge), Université de Paris III.

⁹ CELLARD, Jacques (1979) : *La vie du langage*, coll. «L'ordre des mots», Paris. Le Robert, p.9.

la presse française¹⁰ (dont en autres *L'Express*) et dans deux titres de la presse britannique¹¹.

Mis à jour deux fois par mois, ce serveur présente l'avantage de posséder un moteur de recherche qui permet de sélectionner une offre selon plusieurs critères : fonction (*direction, production, commercial...*) secteur (*BTP, santé, tourisme, mécanique, énergie, chimie...*) lieux (*région parisienne, province...*). Mais surtout il existe aussi une possibilité de recherche par mots-clés et c'est à partir de celle-ci que nous avons poursuivi notre étude.

Nous avons retenu dans un premier temps tous les emprunts de notre liste totalisant un nombre d'occurrences supérieur à 5 et susceptibles par ailleurs de constituer un mot-clé afin de déterminer ceux qui étaient reconnus par le serveur. C'est ainsi que nous avons pu lancer une recherche à partir des termes suivants :

<i>marketing</i> (129)	<i>high tech</i> (60)
<i>leader</i> (401)	<i>merchandiser</i> (1)
<i>challenge</i> (65)	<i>prospect</i> (10)
<i>manager</i> (71)	<i>reporting</i> (16)
<i>merchandising</i> (12)	<i>holding</i> (15)
<i>packaging</i> (8)	<i>mix</i> (4)
<i>business to business</i> (12)	<i>process</i> (18)
<i>leadership</i> (26)	

Les chiffres qui figurent entre parenthèses sont les nombres d'offres correspondant à la demande et affichés par le serveur en fonction du mot-clé utilisé. Les chiffres donnés ici sont ceux obtenus lors d'une recherche sur «Cadres on line» entre le 13 juin et le 3 juillet 1997 à partir de 2043 offres d'emploi. Nous ne les communiquons qu'à titre indicatif et il ne peuvent pas être rapprochés du nombre d'occurrences signalé dans notre liste d'emprunts lexicaux (Annexe A) car chaque mot, même répété plusieurs fois dans une annonce, n'est compté qu'une seule fois par le système de recherche.

En revanche nous avons veillé à ne pas prendre en compte les offres rédigées en anglais et signalées par le serveur à l'aide d'un petit drapeau britannique.

Nous avons ensuite effectué une recherche en utilisant cette fois comme mots-clés, non plus les termes anglo-américains cités plus haut mais au contraire les néologismes dont l'usage est fortement encouragé par les pouvoirs publics.

Comme nous le présentions, les résultats ont été extrêmement décevants.

En ce qui concerne les néologismes suivants :

mercatique (= *marketing*)
chalenge (= *challenge*)

¹⁰ *L'Express, Le Monde, L'usine nouvelle, 01 Informatique, Electronique International, 01 Réseaux, Internet professionnel, Décision Micro et Réseaux, Caractère, France agricole, Agro Distribution, RIA, Néo-Restoration, LSA, L'Echo, Le moniteur, L'Argus.*

¹¹ *Property week, Building.*

manageur (= *manager*)
marchandisage (= *merchandising*)
marchandiseur (= *merchandiser*)
mercatique après-vente (= *reporting*)
tenante (= *holding*)
marchéage (= *mix*)

le serveur nous a répondu à chaque tentative de recherche :

«Nous sommes désolés. Aucune offre ne correspond à votre demande».

(Notons au passage qu'une simple modification orthographique du terme emprunté visant à le mettre en conformité avec notre système morpho-phonologique (*manager* = *manageur*, *challenge* = *chalenge*) suffit à faire obstacle à son intégration.)

Même réponse négative à partir de la périphrase descriptive :

négociation d'entreprise à entreprise (*business to business*)

En revanche, lorsque nous avons tenté une recherche portant sur un essai de traduction du terme d'emprunt, nous avons enregistré quelques réponses positives en particulier pour :

conditionnement (9) qui obtient un résultat légèrement supérieur à *packaging*

de même que *procédé* (7) qui pourrait se substituer à *process*.

Mais comme nous le verrons plus tard la traduction d'un emprunt par un terme univoque n'est pas toujours possible. Ainsi une recherche menée sur les mots *meneur* et *autorité* en traduction de *leader* et *leadership* renvoie aux qualités requises chez un candidat et ne rend pas compte de la notion de position dominante occupée par une entreprise sur un secteur donné.

Conforté, au terme de cette expérience télématique, dans notre opinion concernant la très faible diffusion des néologismes dans le monde professionnel, nous avons essayé d'analyser les différents facteurs pouvant expliquer une telle indifférence à l'égard de la terminologie officielle.

UNE INFORMATION SUFFISANTE

D'emblée nous avons éliminé l'argument selon lequel ce phénomène relève d'un simple manque d'information.

En effet les travaux des commissions ministérielles de terminologie ont toujours fait l'objet d'une très large diffusion auprès des professionnels et du grand public, soit par le biais d'organes officiels (Les Notes Bleues de Bercy), soit par la presse grand public (*Le Monde*, *Libération*...), soit encore par des organismes privés tels que l'APFA (Association pour Promouvoir le Français des Affaires), association à but non lucratif qui organise chaque année sous le patronage du Commissariat général de la langue française «La journée du français des affaires» et qui publie sous forme de dépliant «Les 700 mots d'aujourd'hui pour les affaires».

TRADUCTIONS TARDIVES

En fait, la première raison qui vient naturellement à l'esprit pour expliquer les réticences des professionnels à utiliser les néologismes officiels se situe dans le décalage que l'on enregistre entre le moment où un terme anglo-saxon est introduit dans notre langue et le moment où un équivalent français est enfin proposé.

L'exemple le plus connu est bien sûr celui de *marketing/mercatique*.

Introduit en France dans les années 1950, le concept nouveau de *marketing* est resté sans traduction jusqu'en 1973, date à laquelle l'emploi de *mercatique* fut d'abord conseillé, puis rendu obligatoire en 1987. *Mercatique* et ses dérivés *mercaticien(-ne)* se proposent de traduire la multitude d'appellations apparues anarchiquement dans les entreprises (*marketing man, marketing manager, marketing management...*) mais, ainsi que nous avons pu le constater dans notre étude, ils demeurent complètement inconnus des professionnels.

Ce décalage est encore accentué par l'évolution très rapide des sciences et des techniques. Les terminologues ont de plus en plus de mal à coller à l'événement linguistique ainsi que le rappelle Claude Hagège : «...le rythme des apports étrangers, accrédités par les médias, est aujourd'hui trop fébrile pour laisser place aux créations lexicales qui arrivent longtemps après la chose.»¹²

Sans doute est-ce ce retard qui permet d'expliquer d'autres échecs comme *marchéage* (*marketing mix*), *marchandisage* (*merchandising*), *minimarge* (*discount*), *minimargeur* (*discounter*), *stylique* (*design*)...

Comment fait-on aujourd'hui pour diffuser les équivalents de *trade marketing* (mercatique de la distribution) et de *reporting* (mercatique après-vente) alors que *mercatique* est lui-même ignoré ?

Le *management* décrit dès 1918 par Henri Fayol¹³, a été officiellement francisé en 1973. Mais combien de temps faudra-t-il encore pour diffuser les équivalents de ses dérivés les plus récents : *disease management* (prise en charge globale de la maladie), *data management* (gestion des données), *facilities management* (gestion des infrastructures) ?

TRADUCTIONS MULTIPLES

À la décharge des utilisateurs d'une terminologie d'emprunt il faut rappeler les hésitations qui caractérisent parfois les directives officielles. Ainsi devant l'échec manifeste d'une première traduction, les commissions de terminologie, après plusieurs années de réflexion, proposent un autre équivalent qui se trouve, pendant quelque temps, en concurrence directe avec le premier terme proposé. On imagine l'embarras du locuteur francophone qui, contraint de faire un choix délicat, optera finalement pour l'expression anglo-saxonne.

¹² HAGÈGE, Claude (1987) : *Le français et les siècles*, Paris, Éditions Odile Jacob, p. 150.

¹³ FAYOL, Henri (1918) : *Administration industrielle et générale*, Paris, Dunod.

Rappelons certaines de ces traductions multiples que l'on a pu rencontrer au cours de ces vingt dernières années :

consumerism fut traduit par *consumérisme* (traduction proposée en 1979 par l'Académie des Sciences Commerciales avec la caution du Conseil International de la Langue Française) alors qu'une circulaire ministérielle du 3 janvier 1994 recommandait l'usage des termes *consommatisme* et *consommaction*;

ou bien

sponsor, traduit par *commanditaire* dans l'arrêté du 17 mars 1982 relatif au vocabulaire du tourisme et par *parrain* et/ou *parraineur* par l'arrêté du 6 janvier 1989 relatif à la terminologie économique et financière;

ou encore

design, qui fut pendant longtemps traduit par *esthétique industrielle* jusqu'à ce que soit inventé le terme *stylique* en 1983.

EMPRUNTS POLYSÉMIQUES

Il arrive parfois qu'un emprunt s'établisse durablement dans sa langue d'accueil lorsqu'il possède un champ sémantique plus large que celui couvert par ses différentes traductions. Ainsi le mot *challenge* (que l'on tente d'adapter en français sous la forme *chalenge* en raison d'une étymologie ancienne et de la règle de doublement des consonnes)¹⁴ ne peut se satisfaire de la traduction *défi*, car ce terme contient également la notion de *remise en question*, d'*entreprise périlleuse* ou encore de *pari*.

Les annonces que nous avons dépouillées en témoignent :

«Homme de challenge vous aimez relever les défis ?» (*L'Express* n° 2390 p. 8)

«Si vous êtes un homme ou une femme de défi et de challenge, rejoignez-nous» (*L'Express* n° 2394 p. 58)

Nous retrouvons le même phénomène avec les mots *leader* et *leadership* que nous avons évoqués précédemment. Empruntés au XIX^e siècle au vocabulaire anglo-saxon de la politique, ces deux termes couramment utilisés aujourd'hui dans les offres d'emploi doivent leur succès à leur multiplicité de sens.

leader : chef, meneur d'homme, chef de file, numéro un, premier, principal..., ex :

«Vous êtes avant tout un leader, capable d'organiser et fédérer votre équipe» (*L'Express* n° 2391 p. 23)

«Nous sommes leader européen de la distribution par catalogue des matériels électroniques...» (*L'Express* n° 2391 p. 28)

¹⁴ VOIROL, Michel (1989) : *Anglicismes et anglomanie*, Paris, Éd. Centre de formation et de perfectionnement des journalistes, p. 19.

«Vous prenez en charge le produit leader du groupe.» (*L'Express* n° 2391 p. 35)

leadership : autorité, sens du commandement, place de premier, position dominante..., ex :

«Excellent communicateur, faisant preuve de leadership, vos qualités d'organisateur et de coordinateur ne sont plus à démontrer.» (*L'Express* n° 2392 p. 20)

«Nous vous offrons :...Un leadership incontesté, des marques de forte notoriété...» (*L'Express* n° 2390 p. 27)

C'est encore la polysémie qui peut justifier l'emprunt du terme *business* tel qu'il apparaît dans les offres d'emploi :

«Vous gérez votre business et pilotez ce centre de profit avec une réelle autonomie» (*L'Express* n° 2396 p. 38)

Traduit en français *business* oblige à faire un choix entre *les affaires, l'affaire, l'entreprise, le commerce, la clientèle* ou *le secteur d'activité...*

EMPRUNTS PLUS COURTS

On ne peut parler de l'anglicisation du français sans évoquer le fait que les emprunts à l'anglais sont souvent beaucoup plus courts, et donc moins coûteux à produire que leurs équivalents français. Ce détail n'a pourtant pas échappé aux terminologues qui se sont efforcés de créer des termes au moins aussi courts que les emprunts qu'ils souhaitaient remplacer. *Mercatique, informatique, ingénierie, stylisme, tenante*, en sont quelques exemples. Malheureusement la structure de notre langue nous force à des créations plus longues.

Quelles chances, de ce point de vue, avons-nous de voir le mot *reporting* être remplacé par *mercatique après-vente* (domaine mercatique), *déclaration des ordres* (domaine bancaire) ou encore *reddition des comptes* (domaine comptable) ? Il en est de même pour *démarchage téléphonique* (phoning), *publipostage* (mailing), *réunion préparatoire* (briefing) ou la *mercatique de la distribution* (trade marketing) parfois nommée *distrimercatique*¹⁵.

EMPRUNTS CONNOTATIFS

Les arguments que nous venons d'évoquer permettent de justifier la nécessité matérielle d'utiliser dans une langue professionnelle une terminologie empruntée à l'anglo-américain. Ces arguments, qui ne prennent en compte que la fonction dénotative des emprunts lexicaux, peuvent être acceptés dans la perspective d'une langue technique, fonctionnelle et univoque.

Cependant, qui pourrait croire que la langue des offres d'emploi est technique au point d'être réductible à sa seule fonction référentielle ? Outre le fait qu'il n'existe pas, dans les langues naturelles, de terme purement dénotatif, il est facile de démontrer que

¹⁵ MOULINIER, René (1997) : *Les 500 mots-clés de la vente*, Paris, Dunod, p. 351.

chaque terme d'emprunt est toujours plus ou moins utilisé en fonction de sa valeur connotative

Prenons le terme *manager*. Ce terme est en anglais relativement neutre dans la mesure où il peut s'appliquer indifféremment à un directeur de société ou à un gérant de station-service. Or, si cet emprunt prolifère dans les offres d'emploi, c'est essentiellement en raison de ses connotations. Il désigne ainsi une personne occupant un poste de responsabilités qui adopte nécessairement des méthodes de gestion modernes importées des États-Unis. Naturellement la modernité, l'efficacité voire la jeunesse ne sont connotées que dans la mesure où le terme est perçu comme américain. Le travestir dans la graphie francisée *manageur* lui ôte toute sa puissance évocatrice et le rabaisse au rang de nos *responsables, directeurs* et autres *chefs de service*. Les entreprises ne s'y trompent pas et usent de cet artifice pour le recrutement de leurs cadres :

«Chef des ventes.....grâce à vos qualités de manager...» (*L'Express* n° 2394 p. 37)

«Chefs de secteur (managers commerciaux)» (*L'Express* n° 2394 p. 37)

«Technical Sales Manager (responsable de zone)» (*L'Express* n° 2394 p. 40)

ou encore cette définition qui laisserait rêveur un lecteur anglophone :

«Meneur d'hommes, Entrepreneur, en un mot Manager» (*L'Express* n° 2394 p. 32)

Dans bon nombre de cas, il est évident que la valeur connotative de l'emprunt constitue sa seule justification. Nous le constatons lorsque l'emprunt n'est pas introduit pour désigner une réalité anglo-saxonne encore mal connue ou mal acclimatée en France, mais au contraire pour exprimer un concept parfaitement français et pour lequel notre langue possède sa propre désignation.

Que dire en effet des termes *packaging* (conditionnement), *sales* (ventes), *discount* (rabais), *consumer* (consommateur), *retail* (détail) ou encore *process* (procédé) ?

Notons à ce propos que lorsque l'emprunt n'est pas motivé par un réel besoin dénotatif il est soumis aux caprices de la mode et sa durée de vie en français peut être très variable. Ainsi les termes *know-how*, *team*, *winner*, *deal*, très courants dans les années 1980 ont aujourd'hui presque tous disparu.

Même si la majorité des emprunts ont été à l'origine introduits pour palier une lacune linguistique réelle, la plupart se maintiennent dans notre langue par la seule force de leurs connotations. Et c'est en fait un véritable discours qui se met en place sous l'alibi d'une langue technique. Un discours très idéologique qui n'est autre que celui du culte de l'Amérique comme modèle de réussite économique et commerciale. Dénoncée dans les années 1960, «l'américanophilie» a été relayée depuis cette époque par les cadres français de l'industrie et du commerce qui adoptent avec enthousiasme un vocabulaire d'outre-Atlantique, dans le seul but de faire rejaillir sur notre réalité économique hexagonale une partie du prestige du modèle américain.

Les terminologues ont beau s'acharner, ni la mercatique, ni le marchandisage n'évoqueront le rêve américain et les entreprises continueront de recruter des *sales managers*...

Naturellement, la maîtrise d'un vocabulaire pseudo-technique joue un rôle social important dans la mesure où elle renforce l'image d'un groupe d'initiés.

Comme le soulignait Claude Hagège : «Les anglicismes techniques tiennent le profane à distance. Ils flattent celui qu'ils distinguent»¹⁶. Mais cette recherche de la distinction à tout prix marque aussi les limites du discours technique. Perçu par le grand public comme une garantie de savoir, l'usage injustifié d'une terminologie d'emprunt n'est pas sans évoquer le latin de cuisine des médecins de Molière et témoigne souvent d'une indigence de la pensée. En effet, si l'emprunt permet de faire l'économie d'une traduction difficile, il dispense également son utilisateur d'une indispensable réflexion à propos de l'objet dénoté. «Le connoté prend alors le pas sur le dénoté et le terme d'emprunt perd toute sa justification du point de vue d'un discours qui se prétend, avant tout, technique et monoréférentiel»¹⁷.

CONCLUSION

Cette étude sur la terminologie en usage dans les offres d'emploi nous a tout d'abord permis de relativiser l'importance de l'emprunt lexical dans ce domaine très circonscrit de la langue des affaires. Loin du raz de marée qu'une lecture un peu rapide de la presse pour cadres pourrait laisser envisager, le phénomène se limite à la généralisation de deux termes (*marketing* et *leader*) qui selon toute vraisemblance s'imposeront définitivement dans notre langue avec leur graphie originale, comme se sont autrefois imposés *week-end*, *gadget* ou *hold-up*. L'emploi des autres termes dépend plus de la nouveauté qu'ils représentent et des connotations qui leur sont attachées, que d'une véritable nécessité dénomminative.

En revanche il faut constater une réelle résistance de la part des milieux professionnels à utiliser la terminologie officielle mise en place par les pouvoirs publics. Certes une langue ne se gouverne pas par décrets et son usage réel s'établit au niveau de toute une communauté dont les réactions peuvent être orientées mais jamais imposées.

On pourrait toutefois déplorer la méfiance collective de la communauté francophone de France à l'égard de toute activité néologique. Les vains efforts des femmes ministres du nouveau gouvernement de Lionel Jospin pour imposer l'expression «Madame la Ministre» en sont un très bel exemple.

L'héritage de trois cents ans de purisme langagier, et ce n'est pas le moindre des paradoxes, a hérissé notre langue de tant d'interdits que les francophones ainsi «privés de leur dynamisme inventif»¹⁸ préférèrent emprunter un terme étranger qu'utiliser une expression nouvelle qu'ils jugent, a priori, suspecte.

¹⁶ HAGÈGE, Claude : *op.cit.* p. 115.

¹⁷ SOUBRIER, Jean (1985) : *Le français économique et commercial : Ambiguïté d'une langue parallèle*, Thèse de Doctorat, Université Lyon 2, p. 174.

¹⁸ HAGÈGE, Claude : *op.cit.* p. 172.

Annexe A1

Emprunts lexicaux - Classement par occurrences

marketing	552	géomarketing	2
leader	425	hardware	2
challenge	96	key account	2
manager	61	Key Account Manager	2
merchandising	26	marketing mix	2
packaging	23	marketing on line	2
business to business	15	mediaplanning	2
leadership	15	outdoor	2
high tech	12	project director	2
merchandiser	11	screening	2
prospect	11	show room	2
reporting	9	success story	2
holding	7	web	2
mix	7	back office	1
process	6	biomedical operation manager	1
B to B	5	brief	1
design	5	business analyst	1
helpdesk	5	business plan	1
sportswear	5	business school	1
British	4	business unit	1
business	4	business unit manager	1
corporate	4	cash and carry	1
data management	4	challenger	1
engineering	4	co-leader	1
laser	4	coaching	1
mix marketing	4	commando	1
start up	4	consulting	1
télémarketing	4	consumer marketing manager	1
trade marketing	4	credit manager	1
broadcast	3	data communication	1
CD-Rom	3	data manager	1
internet	3	digital mapping	1
mailing	3	direct marketing	1
medico-marketing	3	drug master files	1
mixt	3	drug safety officer	1
monitoring	3	entertainment	1
planning	3	entrepreneurial	1
pressing	3	Europe's first marketing agency	1
pricing	3	European Key Accounts	1
scanner	3	facilities management	1
business development	2	fax	1
cocktail	2	fluent	1
consumer	2	fluid handling components	1
data	2	full service	1
discount	2	fun	1
disease management	2	get yourself connected	1

hard	1
healthcare	1
incentive	1
interviewer	1
intranet	1
job	1
kit	1
labelling	1
laser disc	1
leasing	1
merchandising	1
marketing B to B	1
marketing promotion	1
meeting	1
micro-marketing	1
mini disc	1
new biz	1
office product	1
on line	1
on time delivery	1
organizer	1
outplacement	1
outsourcing	1
package carrier	1
package express	1
pool	1
product line management	1
provider	1
purchasing manager	1
push technology	1
recording	1
remodeling	1
retail	1
retail marketing manager	1
royalties	1
sales and marketing	1
sales engineer	1
sales manager	1
sales operation	1
search	1
senior management program	1
soft drink	1
sourcing	1
stock option	1
stress	1
team leader	1
technical sales manager	1
test	1
top	1
total look	1

Annexe A2		Emprunts lexicaux - Classement alphabétique	
B to B	5	full service	1
back office	1	fun	1
biomedical operation manager	1	géomarketing	2
brief	1	get yourself connected	1
British	4	hard	1
broadcast	3	hardware	2
business	4	healthcare	1
business analyst	1	helpdesk	5
business development	2	high tech	12
business plan	1	holding	7
business school	1	incentive	1
business to business	15	internet	3
business unit	1	interviewer	1
business unit manager	1	intranet	1
cash and carry	1	job	1
CD-Rom	3	key account	2
challenge	96	Key Account Manager	2
challenger	1	kit	1
co-leader	1	labelling	1
coaching	1	laser	4
cocktail	2	laser disc	1
commando	1	leader	25
consulting	1	leadership	15
consumer	2	leasing	1
consumer marketing manager	1	mailing	3
corporate	4	manager	61
credit manager	1	marchandising	1
data	2	marketing	552
data communication	1	marketing B to B	1
data management	4	marketing mix	2
data manager	1	marketing on line	2
design	5	marketing promotion	1
digital mapping	1	mediaplanning	2
direct marketing	1	medico-marketing	3
discount	2	meeting	1
disease management	2	merchandiser	11
drug master files	1	merchandising	26
drug safety officer	1	micro-marketing	1
engineering	4	mini disc	1
entertainment	1	mix	7
entrepreneurial	1	mix marketing	4
Europe's first marketing agency	1	mixt	3
European Key Accounts	1	monitoring	3
facilities management	1	new biz	1
fax	1	office product	1
fluent	1	on line	1
fluid handling components	1	on time delivery	1

Néologismes et termes d'emprunt dans le français des affaires

organizer	1	sales and marketing	1
outdoor	2	sales engineer	1
outplacement	1	sales manager	1
outsourcing	1	sales operation	1
package carrier	1	scanner	3
package express	1	screening	2
packaging	23	search	1
planning	3	senior management program	1
pool	1	show room	2
pressing	3	soft drink	1
pricing	3	sourcing	1
process	6	sportswear	5
product line management	1	start up	4
project director	2	stock option	1
prospect	11	stress	1
provider	1	success story	2
purchasing manager	1	team leader	1
push technology	1	technical sales manager	1
recording	1	télémarketing	4
remodeling	1	test	1
reporting	9	top	1
retail	1	total look	1
retail marketing manager	1	trade marketing	4
royalties	1	web	2

Annexe B

Emprunts sémantiques - Classement par occurrences

approche	manière d'aborder (un problème)	4
attractif	attirant	31
auditeur	vérificateur, commissaires aux comptes	3
contrôle	maîtrise, direction	4
contrôler	diriger	2
developper	accroître	12
developpeur	chargé d'accroître (un volume d'affaires)	10
en charge de	responsable de	9
junior	débutant	31
major (adj.)	principal	1
opportunité	occasion	80
rapportant (part. pré)	(to report to) : rendre compte à	3
reportant (part.pré)	(to report to) : dépendre hiérarchiquement	2
senior (adj.)	confirmé	51

Emprunts sémantiques - Classement par occurrences

opportunité	occasion	80
senior (adj.)	confirmé	
attractif	attirant	31
junior	débutant	
Développer	accroître	
développeur	chargé d'accroître (un volume d'affaires)	10
en charge de	responsable de	9
approche	manière d'aborder (un problème)	4
contrôle	maîtrise, direction	4
auditeur	vérificateur, commissaires aux comptes	3
rapportant (part. pré)	(to report to) : rendre compte à	3
contrôler	diriger	
reportant (part.pré)	(to report to) : dépendre hiérarchiquement	2
major (adj.)	principal	

ENTRE STOCKAGE ET RÉEMPLOI EN TERMINOLOGIE : UN PROCESSUS D'ANAMNÈSE ?

Michèle A. LORGNET

Université de Bologne-Forlì, Italie

On présente ici une hypothèse de recherche qui a pour origine une réflexion sur l'utilisation de la terminologie dans ses deux aspects et fonctions : le stockage des connaissances et leur réemploi successif en traduction vers la langue étrangère. La recherche emprunte une méthode empirique analysant le double apprentissage des moyens terminologiques et terminographiques et de leur utilisation en traduction professionnelle pour de jeunes apprenants-traducteurs.

Les apprenants concernés, les étudiants en traduction et interprétation de l'Université de Bologne-Forlì (SSLiMIT), ont été observés pendant la période de rédaction de leur mémoire de maîtrise. Pour vérifier, ou anticiper, durant cette période, la capacité professionnelle de ces futurs traducteurs *in situ*, (Gouadec 1994 : 163-182), c'est-à-dire dans des entreprises et des organismes de la Romagne, nous avons constitué un groupe de travail de chercheurs et d'enseignants pour assister ces «impétrants» en traductologie, et ce pour toutes les langues enseignées à l'école, c'est-à-dire l'anglais, le français, l'allemand, l'espagnol et le russe. Une période de stage et de collecte de matériel à l'intérieur des entreprises a donc été doublée et suivie d'une période d'informatisation assistée.

Dans les pays de langue minoritaire comme l'Italie, la plupart des travaux de traductions techniques requis par les petites et moyennes entreprises sont des traductions vers la langue étrangère. Il s'agit dans la grande majorité des cas de textes pour la présentation ou l'entretien de matériel, textes allant du simple opuscule au livret technique élaboré, du simple feuillet illustratif accompagnant la machine ou le produit, au volume présenté en librairie, par exemple, en ce qui concerne un des organismes concernés. L'hypothèse de recherche collective a donc envisagé plusieurs stades dans l'élaboration pédagogique nécessaire à une telle assistance. Pour pouvoir développer «l'autonomie de production de traduction» des étudiants, il a d'abord fallu leur demander d'informatiser un matériel linguistique à l'état brut, concernant une terminologie extrêmement variée, difficile à classer, en micro-glossaires bi- ou trilingues. Pour tester leur capacité de réemploi du matériel terminologique informatisé par leurs soins lors de traductions vers la langue étrangère à partir de l'italien, il a été ensuite nécessaire de contrôler les micro-corpus en corpus parallèles (Lorgnet 1997 : 7-9).

La période de stages financés par la municipalité, la Chambre de Commerce et les banques de Forlì, avait garanti le côté officiel de cette première entrée dans le monde du

travail. Les étudiants effectuaient ainsi la période de collecte de matériel à l'intérieur des entreprises, travaillant généralement par groupe de deux, pour un maximum de trois cents heures de recherche. Ces étudiants étaient tous inscrits à la dernière année de leurs cursus universitaires. En même temps, ils constituaient les fiches terminologiques bi- ou trilingues des micro-glossaires dans une série de domaines qui étaient, par conséquent, en même temps extrêmement restreints au niveau de la spécialisation et extrêmement larges et flous au niveau de la nature des textes traduits par la suite.

Les domaines concernés étaient les suivants : artisanat, agriculture intégrée, industrie mécanique — élévateurs hydroliques, chaudières à gaz, agraieuse, machines agricoles pour jardinage —, biologie, préservation de l'environnement d'un parc national, le Parc National du «Casentino» ayant, à l'occasion, demandé la traduction d'un important volume concernant l'histoire, la géologie, la faune et la flore de ce milieu naturel. La constitution des fiches de données était toujours soumise à un triple contrôle : contrôle linguistique des enseignants, confrontation avec les documents sur la normalisation des termes à la Chambre de commerce de Forlì, consultation des banques de données existant sur le marché et consultation technique des ingénieurs et spécialistes des entreprises. L'assistance technique du travail informatique a été effectuée par Franco Bertaccini.

La fiabilité du travail terminologique accompli était vérifiée en retour dans des traductions de documents professionnels vers la langue étrangère avec confrontation directe du travail par les partenaires étrangers des entreprises.

À la suite de résultats assez fluctuants dans la recherche, puis ensuite dans l'utilisation des termes, leur classification et le réemploi successifs de notions techniques ou technologiques dans une langue étrangère (Assal 1994 : 461-462), il a fallu essayer d'identifier le pourquoi de ces nombreuses difficultés d'apprentissage.

Tout d'abord, l'abstraction notionnelle nécessaire à l'opération de stockage de la terminologie semblait poser des problèmes dans l'opération successive du réemploi en langue étrangère, surtout lorsque le réemploi advenait en un second temps assez éloigné. «Les résidus sémantiques appartenant à un lexème restent collés à lui quand il devient terme» (Assal, 1994 : 462). Le problème ardu de la sélection des synonymes ou quasi synonymes (Kocourek 1982 : 164-167; Lethuillier 1989 : 443-450, Gross et Clas 1997 : 149-150) d'un domaine spécifique ne semblait devoir être résolu que par l'appel à un parlant de langue maternelle (dans notre cas toujours un enseignant de notre groupe de travail).

La prise de conscience de la globalité du contexte terminographique (De Bessé 1991 : 112-113) apparaissait en réalité plus facile à circonscrire lors du stockage que lors du réemploi. La cohérence notionnelle, la consistance des groupements préférentiels (Thoiront et Béjoint 1989 : 666-667), si précieuses dans la langue maternelle, ne semblait souvent à nos traducteurs en herbe bien plus une difficulté seconde qu'un atout lors de leur travail de mémorisation à long terme. Les imbrications des phénomènes sémantiques avec les connexions syntaxiques (Mel'cuk 1978 : 288-289), particulièrement utiles lors du stockage notionnel, nécessitant toujours une activité contrôlée très forte dans la langue non-maternelle (Gaonac'h 1996 : 27-28), introduisaient quelquefois même un blocage du travail du traducteur vers la langue non maternelle. Le repérage des unités terminologiques dans le contexte de la traduction spécialisée (Mareschal 1989 : 377-378), — effectivement

plus aisé à réaliser sur la base de données textuelles — et la lecture paradigmatique en profondeur du texte à traduire, dans le secteur spécialisé et technique comme dans celui du texte littéraire (Lorgnet 1995), ne garantissaient pas toutefois la même habileté lors de la production en traduction. Après une période d'observation qui n'a somme toute été faite que lors de cette première année de recherche, il nous a semblé que l'opération raisonnée du stockage terme à terme est effectivement partiellement insuffisante pour l'appropriation plus ou moins automatique des termes dans une langue étrangère, appropriation nécessaire lors des traductions rapides effectuées en entreprises. D'autre part, la vitesse de réemploi des syntagmes terminologiques est étroitement liée à la faculté associative des apprenants, même lorsque l'analyse des corpus en parallèle facilite ce réemploi.

Il s'agissait donc de relever, à l'intérieur de ces groupes figés, quels étaient les groupes de mots les mieux stockés, c'est-à-dire qui ne nécessitent plus la consultation des fiches terminologiques en mémoire, pour comprendre si la constitution de réseaux notionnels (Blampain 1992) ou de représentation sémantique est ou non soumise à une mémorisation a-notionnelle fondée sur d'autres facteurs de mémorisation.

La distinction classique entre la capacité de mémorisation à court terme (valeur extensive des termes) et la mémoire à long terme (valeur intensive du terme) semble remise en question lors du réemploi, surtout si l'opération de réemploi intervient sous contrainte temporelle forte. L'instance de contrôle nécessaire à la reverbération en traduction vers la langue étrangère apparaît effectivement n'intervenir qu'en complément des processus d'automatisation (Gaonac'h 1996 : 34). Le travail sur la nature des synonymes en stade de stockage avait utilisé le schéma du CILF pour rendre optimale la compréhension notionnelle des termes. Pour une telle opération, le repérage en langue maternelle du sème culminant sémantiquement est bien sûr très utile (Cigada 1988). Lors du réemploi, on repère malheureusement une sorte de brouillage de la mémoire notionnelle distincte et, lors de la traduction d'un texte d'une certaine longueur, la consultation répétée des fiches en mémoire allonge considérablement le travail.

Exemples

Terme source

terme cible

a) représentation complète en langue source et en langue cible

acclimatamento della pianta

adattamento
compatibilità

acclimatation

adaptation
compatibilité

b) termes équivalents présentant plusieurs synonymes

b1) plusieurs synonymes en langue cible

stolone

selvatico

surgeon

drageon
balais de sorcière

verticillo

verticille

patte d'oie

argilla
mama

argile
marme
glaise
molasse

b2) les termes de la langue source n'ont pas d'équivalent; les synonymes cibles approximatifs ne sont pas équivalents

trattrice agricola

.....

.....
tracteur
véhicule tous terrains

b3) les termes de la langue source n'ont pas d'équivalence synonymique; la langue cible a des synonymes

impagliata

.....

.....
plat à légumes ?

b4) le terme source est acceptable en terme cible

Tatura trellis

Tatura trellis

c1) les termes cibles présentent une traduction oblique

radicelle

(ébauche racinaire)
broussin

c2) le terme cible traduit un synonyme de la langue source

non crescita del meristema

morìa del meristema

ramo spoglio

avortement du méristème
.....
chicot
branche défoliée

Le problème des équivalences sémantiques et de la traduction (Lyadri 1997 : 143) touche donc de très près le rapport avec certains phénomènes linguistiques importants tels que la métonymie (Horacek 1996 : 113) ou si l'on veut la métaphore au sens large et aristotélicien du terme (Lorgnet 1995).

La recherche en cours a donc tenté de repérer quels étaient, dans un domaine sectoriel donné, les syntagmes figés dont la reconnaissance initiale au cours du stockage produit, à long terme, une meilleure production en traduction.

Si l'on part du principe, cher à Halliday, que l'importance du niveau lexicographique l'emporte sur celle du niveau syntactico-sémantique, on peut alors analyser l'efficacité de certains critères comme celui de la boucle phonologique (Baddeley, Papagno, Vallai 1988 : 586-595) et les réseaux associatifs rhétoriques du verbal et du non-verbal.

Avant toute chose, le «style» d'un texte technique ou technologique particulier, c'est-à-dire le surplus de son sens prescriptif et normatif dans une direction-cibliste unilatérale (Dimarco, Mah 1994 : 21-41) se doit d'être perçu et analysé par le traducteur. L'ouverture des choix lexicaux paradigmatiques en langue cible ne semble toutefois perceptible et productive que si la mémorisation syntagmatique s'est, en un certain sens, vidée du sens prescriptif et/ou notionnel. Sans en revenir à un behaviourisme polémique,

il est certain, de par cette première phase de notre recherche, que la mémorisation en série parallèle telle que la représente la mémoire terminologique n'est pas suffisante pour la production en langue étrangère d'un texte cohérent et cohésif. Les «magasins de la mémoire» platonicienne ont d'autres clefs, et il n'est pas si sûr que la mémorisation d'une langue étrangère passe par les mêmes voies que celle de la langue maternelle. Il faudrait à ce propos revoir les assertions et les polémiques sur le bilinguisme précoce.

Dans une série de termes synonymes, la mémorisation la meilleure semble presque toujours, de par notre première expérience, concerner le premier terme de la liste qui suit. Quand la mémorisation manque complètement, la recherche notionnelle des synonymes en langue source, ou bien l'analyse de l'apex sémantique, régénère le terme cible oblique, c'est-à-dire le terme en forme de syntagme ou altéré par une transformation sémantique morphématique ou métaphorique.

Fil d'Archal

fil de laiton

cruchon

pichet
broc

dormance

latence

habiller les racines

pralinage
préparation des racines

bouquet de mai

rosette de bourgeons (de feuilles)

sevrage

taille au collet

balais de sorcière

surgeon

CONCLUSION

Une analyse prolongée et systématique de ces intuitions pourra sans doute être mieux élaborée par la suite, en multipliant les sujets d'observation. Il reste toutefois que l'effort mnémotechnique des apprenants en terminologie pour la traduction vers la langue étrangère fait appel à des mécanismes échappant à un notionalisme pur et dur, et qu'une meilleure prise de conscience de libres associations synonymiques en langue maternelle grâce à des exercices préparatoires en classe de traduction aide les futurs traducteurs à développer une fonction parallèle d'anamnèse également dans la langue étrangère.

RÉFÉRENCES

- ASSAL, A. (1994) : «La notion de notion en terminologie», *META*, XXXIX, 3, pp. 460-464.
- BADDELEY, A.D., PAPAGNO, C., et G. VALLAR (1988) : «When long-term learning depends on short-term storage», *Journal of Memory and Language*, 27, pp. 586-595.
- BLAMPAIN, D. (1992) : «Traduction et écosystèmes terminologiques», *Équivalences*.
- DE BESSE, B. (1991) : «Le contexte terminographique», *META*, XXX (1), pp. 111-120.
- CIGADA, S. (1988) : «I meccanismi del senso : Il culminatore semantico», E. Rigotti et C. Cipolli (dir), *Ricerche di Semantica testuale*, Brescia, La Scuola, pp. 25-70.
- DIMARCO, Ch. et K. MAH (1994) : «A Model of Comparative Stylistics for Machine Translation», *Machine Translation*, pp. 21-60.
- GAONAC'H, D. (1996) : «Processus cognitifs de base dans l'acquisition des langues», *Bulletin suisse de linguistique appliquée*, 63, pp. 25-36.
- GOUADEC, D. (1994) : «Terminologie, phraséologie et informatique dans la formation des traducteurs professionnels», M. Lorgnet (dir), *Atti della Fiera Internazionale della Traduzione, Forum di Forlì, 3-6 dicembre 1992*, «Biblioteca della Scuola Superiore di Lingue Moderne per Interpreti e Traduttori», 4, Bologna, CLUEB, pp. 163-183.
- GROSS, G. et A. CLAS (1997) : «Synonymie, polysémie et classes d'objets», *META*, XLII (2), pp. 147-154.
- HORACEK, H. (1996) : «On expressing Metonymic Relations in Multiple Languages», *Machine Translation*, 11, pp. 109-158.
- KOCOUREK, R. (1982) : *La langue française de la technique et de la science*, Wiesbaden, Branstetter Verlag.
- LETHUILLIER, J. (1989) : «La synonymie en langue de spécialité», *META*, XXXIV (3), pp. 443-449.
- LORGNET, M.A. (1995) : *Pour une traduction holistique. Recueil d'exemples pour l'analyse et la traduction*, Bologna, Clueb, pp. 135.
- LORGNET, M.A. (à paraître) : «Coherence into Non-mother Tongue Terminology», *Proceedings of the Conference "Translation into Non-Mother Tongues"*, Ljubljana, 29-30 May 1997.
- LYADRI, R. (1997) : «Problématique des équivalences sémantiques et de la traduction dans des dictionnaires arabe-français», *META*, XLII (1), pp. 142-146.
- MARESCHAL, G. (1989) : «Repérage d'unités terminologiques dans le contexte de l'enseignement de la traduction spécialisée», *META*, XXXIV (3), pp. 377-380.
- MEL'CUK, I. (1978) : «Théorie de langage, théorie de traduction», *META*, XXIII (4), pp. 271-302.
- THOIRON Ph. et H. BÉJOINT (1989) : «Pour un index évolutif et cumulatif de cooccurrents en langue techno-scientifique sectorielle», *META*, XXXIV (4), pp. 661-671.

INSTITUTIONS LEXICALES ET DICTIONNAIRES TRADITIONNELS MALGACHES*

Jean-Yves MORIN

Université de Montréal, Canada

1. INTRODUCTION

Les *dictionnaires traditionnels*, malgré leurs défauts, ont l'immense avantage d'être aisément manipulables par les humains (et, jusqu'à un certain point, par les machines). Dans les bases lexicales à couverture large, on aimerait pouvoir intégrer ces divers types d'information et générer, de façon automatique ou semi-automatique, des représentations riches et souples. Les premiers travaux portant sur l'exploitation informatique des dictionnaires traditionnels (sous format électronique)¹ avaient des buts à la fois trop ambitieux et trop limités. D'une part, ils voulaient dériver automatiquement une représentation sémantique formelle à partir des définitions contenues dans ces dictionnaires en sous-estimant grandement la complexité et le manque de fiabilité d'une telle transformation. D'autre part, la sémantique lexicale sous-jacente était extrêmement étroite (pour ne pas dire simpliste) et l'on n'utilisait généralement qu'une seule source de données². Ce que nous proposons ici, c'est de reconsidérer le problème sous un autre angle, en tirant le meilleur profit possible des dictionnaires traditionnels. Il ne s'agit pas de proposer un nouveau formalisme de représentation de l'information lexicale, il en existe déjà plusieurs³, ni un ensemble d'algorithmes de restructuration de lexiques⁴, mais plutôt un ensemble de méthodes d'analyse fonctionnelle des données lexicales, les *institutions lexicales*, qui permettent un typage rapide de propriétés en termes de contenus et qui

* Une version préliminaire de ce travail a été présentée aux V^{es} Journées scientifiques du réseau LTT *La mémoire des mots* à Tunis, en septembre 1997. Ce travail a bénéficié de l'aide du réseau LTT à l'équipe «TAO et théories linguistiques» (Anne Abeillé, Philippe Blache, Jean-Yves Morin et Éric Wehrli) ainsi que d'une subvention de recherche du fonds FCAR du Gouvernement du Québec. Pour des raisons éditoriales, le contenu de cette présentation a été scindé en trois parties autonomes : «Lexiques-grammaires et institutions lexicales», «Institutions lexicales et DEC» et le présent article.

¹ Cf. par exemple Boguraev & Briscoe (1988).

² Généralement le dictionnaire dont on se trouvait à disposer sous format électronique.

³ Par exemple DATR (Evans & Gazdar 1996), le lexique génératif (Pustejovsky 1995), INTEX (Silberztein 1993).

⁴ Bien que nous utilisions effectivement de tels algorithmes, à titre expérimental.

facilitent l'identification des *conditions de collage* entre institutions lexicales distinctes en fonction des *domaines* qu'elles recouvrent.

Les dictionnaires traditionnels du malgache, comme ceux des langues sémitiques, sont classés par racines, elles-mêmes classées dans l'ordre alphabétique. Si un tel classement peut poser des obstacles à l'utilisateur non malgachophone, qui a souvent du mal à déterminer sous quelle racine chercher une forme donnée⁵, il est particulièrement utile pour la morphologie computationnelle, dans la mesure où le réseau dérivationnel d'une forme y est présenté de façon synoptique. Au niveau de la création de bases lexicales, cette discipline est également très pratique, puisqu'elle permet de créer de façon systématique, en même temps que la racine, un ensemble de dérivés potentiels qui constitueront autant de fragments d'entrées lexicales à compléter, reliés à la racine par des hyperliens.

Les dictionnaires traditionnels ne comportent évidemment pas de représentation sémantique explicite. La composante sémantique y est représentée par des traductions ou des définitions. Il en est de même pour la composante syntaxique, qui se réduit à l'attribution d'une partie du discours à chaque forme. Cependant, la richesse de l'information morphologique et colocationnelle consignée dans ces dictionnaires permet de construire le squelette d'une base de données lexicales susceptible d'être enrichie. En principe, la construction d'une telle base de données constitue un travail long et délicat (comme tout travail lexicographique) qui demanderait non seulement une connaissance parfaite de la langue, mais en outre l'accès à des corpus étendus. Nous nous contentons d'adopter ici une approche à la fois *opportuniste*, qui tente d'exploiter au maximum les informations rassemblées par les lexicographes traditionnels et de les rendre accessibles par des chemins multiples (qui, eux, ne sont pas traditionnels) et *paresseuse*, qui permet de retarder les tâches difficiles (par exemple, l'analyse sémantique fine). La compilation de l'ensemble des formes (racines ou dérivés) en un module (arbre, trie, automate, etc.) permettant d'indexer l'information lexicale sur toutes les formes (et dans toutes les directions) ne pose aucun problème de principe. Une partie importante de l'information lexicale peut donc être directement accessible à un utilisateur quelconque (qui peut aussi bien être un programme qu'un humain) depuis les formes dérivées (et fléchies) vers les formes de base, mais également par plusieurs autres chemins.

À titre d'exemple, considérons l'entrée *varotra* dans trois dictionnaires traditionnels (deux dictionnaires bilingues malgache-français et un dictionnaire monolingue malgache) : Abinal & Malzac (1888, A&M), Rajemisa-Raolison (1985, RR) et Rajaonarimanana (1995, NR). Voyons comment on pourrait exploiter ces informations en déterminant les *conditions de collage* de ces trois entrées à l'intérieur d'un squelette commun.

⁵ Le *Rakibolana malagasy-italiana/Dizionario italiano malgascio* du père Profita (Profita 1969) est organisé de façon alphabétique tant dans sa partie malgache-italienne que dans la partie italienne-malgache et ne contient pratiquement pas de renvois. Il est intéressant de noter qu'à notre connaissance, un des seuls autres dictionnaires où les formes malgaches soient classées par ordre alphabétique (et donc plus accessibles à qui ne connaît pas bien la morphologie malgache) est le *Rakibolana malagasy* (Rajemisa-Raolison 1985, RR), un dictionnaire monolingue malgache, qui s'adresse donc précisément aux malgachophones.

	VAROTRA , <i>s.</i> Vente, commerce, négoce, trafic, marchandises. (Au passif prend. <i>Amidy</i> . Voy. VIDY).
	Mi—, va. Vendre, faire le commerce, trafiquer. Mivaròta , <i>imp.</i> Fi—, s. Ce qu'on peut vendre, manière de vendre. MPI—, s. Celui qui vend, le marchand, le commerçant.
	Fivaròtana , <i>s.</i> L'action de vendre, de faire le commerce, de trafiquer, le prix, le lieu, la cause. Ivaròtana , <i>rel.</i> Ariary telopolo no nivarotako ny omby mifahiko. J'ai vendu trente piastres le bœuf que j'avais engraisé. Ivaròty , <i>imp.</i> Prend Mampi—, Mifampi—.
	Maha—. Qui peut ou sait vendre.
	Varobàrotra , dupl. dim. de Varotra.
1	Mamola-bàrotra . Chercher à faire baisser, à réduire le prix.
2	Manongom-bàrotra . Voy. SONGONA.
3	Tombom-bàrotra , <i>s.</i> Profit, bénéfice.
4	Varotra ambongàdiny , <i>s.</i> Vente en gros, par pièce.
5	Varotra an-trano tòkana , <i>s.</i> Monopole.
6	Varotra an-tsinjàrany , <i>s.</i> Vente au détail.
7	Varo-bàbo , <i>s.</i> Très bon marché, à vil prix. ⁶
8	Varo-balògana . Voy. VALONGANA.
9	Varo-bavènty , <i>s.</i> Commerce en grand; <i>fig.</i> tout vol où on expose sa vie.
10	Varo-bazàha , <i>s.</i> Vente à prix fixe.
11	Varo-bòba , <i>s.</i> Comme VARO-BABO.
12	Varo-dròa , <i>s.</i> Action de vendre encore ce qu'on a vendu, de promettre à l'un ce qu'on a promis à l'autre.
13	Varo-jabo an-tsena, ka samy mandoka ny azy ho màsaka , <i>prov.</i> Chacun vante sa marchandise, comme les marchands de Jabo au marché.
14	Varo-mahay mody et Varo-mièra , <i>s.</i> Vente d'un article qui peut être rendu si l'acquéreur n'en est pas content; action de prendre à crédit avec obligation de rendre l'objet à défaut du prix.
15	Varo-mahèry , <i>s.</i> Vente forcée.
16	Varo-màika , <i>s.</i> Vente faite à vil prix pour cause d'urgence.
17	Varo-mamindro , <i>s.</i> Achat dont on fait traîner la discussion dans l'espérance de trouver mieux.
18	Varo-mandèha , <i>s.</i> Colportage.
19	Varo-màty , <i>s.</i> Vente définitive..
20	Varo-miàndry , <i>s.</i> Objets qu'on garde longtemps, qui ne se vendent que par circonstance, dont on demande un bon prix.
21	Varotry ny mananotèna , <i>lit.</i> Commerce des veuves; <i>fig.</i> vente faite avec hésitation, dans la crainte de se tromper.
22	Varotr'omby anaty ambiàty , <i>lit.</i> Vente de bœufs cachés dans les ambiaty; <i>fig.</i> vente malhonnête.
23	Varo-tsy azon-dàhy , <i>s.</i> Marché sur lequel on ne s'entend pas, malentendu dans une vente ou un achat.
24	Varo-tsy mifòdy , <i>s. lit.</i> Vente sur laquelle on ne revient pas; <i>fig.</i> la mort.

Figure 1 : Varotra (Abinal & Malzac 1888 : 815)

⁶ La forme *varo-babo*, qui apparaît dans Abinal et Malzac, n'est pas reconnue par nos informateurs, seule la forme *varo-boba* existe pour eux. C'est également la seule forme répertoriée dans les dictionnaires plus récents. Ou bien la forme *varo-babo* s'est perdue au profit de son synonyme et quasi-homonyme (équivalent contrepétiq) *varo-boba*, ou bien il s'agit d'une erreur dans A&M. Cependant *varo-babo* y apparaît aussi bien sous l'entrée *babo* que sous celle de *varotra*, si erreur il y a, elle n'est donc pas accidentelle.

	<p>Varotra a. : Fanakalozan-javatra amin'ny vola, hitadiavantombony : <i>Lasa nanao varotra any an-tsiraka izy.</i></p> <p>♦ Entana amidy : <i>Aza hitsahinao ity varotr'olona.</i></p> <p>•● <i>t.i.f.</i> : Voavarotra / mivarotra, fivarotra, mpivarotra / ivarotana, fivarotana / mampivarotra, mifampivarotra, mahavarotra / varobarotra.</p> <p>• F.-p. :</p>
1	<i>Ady varotra</i> : fifamaliana ataon'ny mpivarotra sy ny mpividy momba ny vidin'entana ka ny mpividy mangataka mba hahenan'ny mpivarotra io vidiny io, ny mpivarotra kosa tsy mety mampihena avy hatrany.
2 (< A&M 1)	<i>/ Mamola-barotra</i> : mampitamby ny mpivarotra ny mpividy mba hampihenany ny vidin'ny entana iadiam-barotra, na mampitamby ny mpividy ny mpivarotra mba hanekeny ny vidiny araka ny tiany hivarotana azy.
3 (< A&M 2)	<i>/ Manongom-barotra</i> : mahita olon-kafa mitady hividy ilay zavatra efa nokasaina hovidina ka manondrotra indray ny vidiny efa nomena.
4	<i>/ Mivaro-tena</i> : mahazo vola amin'ny fijanganjangana.
5	<i>/ Varotra alika</i> : varotra ifanerena.
6 (< A&M 4)	<i>/ Varotra ambongadiny</i> : varotra be no atao fa tsy mitsipotipotika.
7	<i>/ Varotra an'elakela-trano, varotra an-kobohobo</i> : varotra atao any antakontakona tsy sahy mazava noho ny mpivarotra ta hanondrotra ny vidin'entana tsy ara-drariny.
8 (< A&M 5)	<i>/ Varotra an-trano tokana</i> : varotra entana tsy mety hita any an-kafa fa ao amin'io toerana io ihany.
9 (< A&M 6)	<i>/ Varotra antsingarany</i> : fivarotana zavatra tsipotehina madinidinka.
10	<i>/ Varotra olona</i> : fivarotana olona atao mpanompo.
11 (< A&M 22)	<i>/ Varotra omby anaty ambiaty</i> : varo-javatra tsy eo ambany maso.
12 (< A&M 21)	<i>/ Varotry ny mananotena</i> : varotry ny mpivarotra misalasala noho izy sady ta hahazo vola no toa tsy mahafoy entana.
13 (< A&M 10)	<i>/ Varo-bazaha</i> : varotra tsy misy ady varotra fa efa raikitra hatrany tsy azo ovana ny vidin'entana.
14 (< A&M 11)	<i>/ Varo-boba</i> : varotra mora dia mora
15	<i>/ Varo-dratsy</i> : fivarotana zavatra nangalarina.
16 (< A&M 12)	<i>/ Varodroa</i> : enti-milaza olona roa samy mivarotra zavatra iray (oh. trano iray izay karakaraina mpanera roa samy hafa).
17 (< A&M 14)	<i>/ Varotra mahay mody</i> : varotra azo averina na afody.
18 (< A&M 16)	<i>/ Varo-maika</i> na <i>Varotry ny maika hiala, Varotry ny maika hody</i> : varo-mora.
19	<i>/ Varo-maizina</i> : mitovy amin'ny varotra an'elakela-trano.
20 (< A&M 18)	<i>/ Varo-mandeha</i> : fivarotana entana mifindrafindra amin'ny toerana maro samihafa fa tsy amin'ny toerana anankiray raikitra.
21 (< A&M 19)	<i>/ Varo-maty</i> : varotra entana tsy azo averina intsony na manao ahoana na manao ahoana. ♦ h.o. : Fahafatesana.
22 (< A&M 20)	<i>/ Varo-miandry</i> : varotra entana izay avela hijanona eo raha mbola tsy lafo.
23	<i>/ Varo-miera</i> na <i>Varo-mifody</i> : varotra entana azon'ny mpividy averina raha tsy mahafa-po azy.
24	<i>/ Varotra misy fehi-vavany</i> : varotra misy fepetra izay azo ravana.
25	<i>/ Varotra tsy miverimbodiondry</i> : varotra entana nanomezana vola mialoha ho fandraiketam-barotra, ka raha tahiny tsy raikitra ny varotra any aoriana, dia lany maina iny vola iny fa tsy averin'ilay olona tompon'entana intsony.

26 (< A&M 24) / *Varo-tsy mifody* : varotra entana tsy azo averina intsony. ♦ h.o. : Fahafatesana.

Figure 2 : Varotra (Rajemisa-Raolison 1985 : 1017)

Voici une traduction approximative de cet article (avec les indices pour faciliter les références)⁷.

«**Varotra** a. : Commerce, vente. Échange d'objet contre de l'argent, dans le but est d'en retirer un profit : *Lasa nanao varotra any an-tsiraka izy. Il est parti faire du commerce dans les régions côtières.* ♦ Marchandise à vendre : *Aza hitsahinao ity varotr'olona. Ne piétine pas ces marchandises.*

•,• Mots de même racine : **Voavarotra / mivarotra, fivarotra, mpivarotra / ivarotana fivarotana / mampivarotra, mifampivarotra, mahavarotra / varobarotra.**

• Expressions : (1) *Ady varotra* : Marchandage. Débat entre vendeur et acheteur sur le prix de la marchandise où l'acheteur demande au vendeur de réduire son prix et le vendeur essaie de baisser petit à petit le prix initial. (2) / *Mamola-barotra* : Négociation sur le prix. L'acheteur menace le vendeur de partir s'il ne veut pas diminuer le prix de sa marchandise ou c'est le vendeur qui fait croire à l'autre que le prix qu'il propose est une bonne aubaine et qu'il doit l'accepter. (3) / *Manongom-barotra* : Enchérissement. Le vendeur augmente le prix sur lequel on s'est mis d'accord parce qu'il a trouvé un autre client qui pourrait acheter la marchandise à un prix plus intéressant pour lui. (4) / *Mivaro-tena* : Se prostituer. Gagner de l'argent avec son activité sexuelle. (5) / *Varotra alika* : litt. "vente de chien" Vente forcée. (6) / *Varotra ambongadiny* : Commerce de gros. Commerce où l'on vend les choses par gros lots, mais en petites quantités. (7) / *Varotra an'elakela-trano, varotra ankobohobo* : litt. "vente dans les petites ruelles", "vente derrière les maisons". Marché noir. Vente illicite pratiquée dans un endroit caché par un vendeur voulant vendre ses marchandises à un prix déraisonnable. (8) / *Varotra an-trano tokana* : Monopole. Commerce où l'on vend des choses qu'on ne peut pas trouver ailleurs. (9) / *Varotra antsingarany* : Commerce de détail. Vente de choses en petites quantités. (10) / *Varotra olona* : Traite d'esclaves. Vente de personnes destinées à l'esclavage. (11) / *Varotra omby anaty ambiaty* : litt. "vente de boeufs dans les ambiaty" Vente d'une chose qui n'est même pas visible. (12) / *Varotry ny mananotena* : Vente où c'est le vendeur qui hésite à savoir s'il peut empocher l'argent tout en gardant la marchandise. (13) / *Varo-bazaha* : litt. "vente à l'étrangère" Vente à prix fixe. Vente sans marchandage. Le prix est fixe et interchangeable. (14) / *Varo-boba* : Liquidation. Vente où les prix sont très bas. (15) / *Varo-dratsy* : litt. "vente méchante". Vente d'objets volés. (16) / *Varodrou* : Vente où deux personnes essaient de vendre la même chose. (Par exemple, deux agents immobiliers qui essaient de vendre une même maison.) (17) / *Varotra mahay mody* : litt. "vente connaissant le retour". Vente où l'on peut échanger ou retourner et se faire rembourser les marchandises. (18) / *Varo-maika na Varotry ny maika hiala, Varotry ny maika hody* : litt. "vente rapide" ou "vente des gens pressés (de rentrer)" Liquidation, vente à bon marché. (19) / *Varo-maizina* : Marché noir. Équivalant à l'expression "varotra an'elakela-trano". (20) / *Varo-mandeha* : Commerce ambulante, colportage. Commerce de marchandises qui se fait en se déplaçant d'un endroit à l'autre, et non dans un seul et même endroit. (21) / *Varo-maty* : Vente définitive. Vente de marchandises qui ne peuvent être ni échangées, ni remboursées, et cela peu importe la raison. ♦ Figuré : la mort. (22) / *Varo-miandry* : litt. "vente dormante" Vente où les marchandises doivent être exposées tant qu'elles ne sont pas

⁷ Cette traduction est présentée suivant la typographie originale de RR, avec l'entrée lexicale en blocs compacts.

vendues. (23) / *Varo-miera* na *Varo-mifody* : Vente à l'essai. Commerce dont la politique consiste à rembourser les marchandises non satisfaisantes (24) / *Varotra misy fehi-vavany* : litt. "vente à bouche cousue" Vente conditionnée par des règles modifiables. (25) / *Varotra tsy miverimbodiondry* : Vente sans retour des arrhes (litt. "vente sans retour du derrière de mouton"). Transaction où l'acheteur doit donner une certaine somme d'avance pour garantir qu'il s'engage à acheter la marchandise mise de côté. Dans le cas où l'acheteur ne voudrait plus prendre la marchandise, son argent ne lui serait pas remis. (26) / *Varo-tsy mifody* : Vente définitive. Vente où les marchandises ne peuvent être retournées. ♦
Figuré : la mort.»

	vàrotra n.	Vente, commerce, marchandises.
1 (< A&M 18, RR 20)	<i>Varo-mandeha,</i>	colportage.
2	<i>Tranom-barotra.</i>	maison, société de commerce.
3 (< RR 1)	<i>Ady varotra,</i>	marchandage, négociation.
4	<i>Fifanakalozam-barotra,</i>	échanges commerciaux.
5	<i>Miady varotra,</i>	marchander, négociier.
6 (< A&M 3)	<i>Tombom-barotra,</i>	bénéfice.
7 (< A&M 24, RR 26)	<i>Varotsimifody,</i>	litt. "vente sur laquelle on ne revient pas", l'au-delà, l'autre monde.
8 (< A&M 11, RR 14)	<i>Varo-boba.</i>	vente à vil prix, solde.
9	<i>Varotra ivelany.</i>	commerce extérieur.
10 (< A&M 19, RR 21)	<i>Varo-maty,</i>	vente définitive.
11 (< A&M 16, RR 18)	<i>Varo-maika,</i>	vente d'urgence.
12 (< RR 19)	<i>Varo-maizina,</i>	marché noir.
13 (< A&M 4, RR 6)	<i>Varotra ambongadiny,</i>	vente en gros.
14 (< A&M 6, RR 9)	<i>Varotra antsinarany,</i>	vente au détail.
15	<i>Vodiondrim-barotra,</i>	arrhes.
16	<i>Varotra iombonana,</i>	communauté économique.
17	<i>Varotra fampisehoana,</i>	exposition-vente.
18	<i>Lalam-barotra.</i>	débouché.
	♦ <i>Mi-</i> , v. a.	Vendre, trafiquer.
	♦ <i>Mpivarotra.</i>	commerçant, marchand.
19	<i>Mpivaro-kena,</i>	boucher.
20	<i>Mpivaro-tena,</i>	prostituée.
21	<i>Mpivarotra amoron-dalana,</i>	marchand du bord des rues.
22	<i>Mpivaro-mandeha,</i>	marchand ambulante.
23	<i>Mpivarotra omby,</i>	marchand de bestiaux.
	♦ <i>Fivarotana,</i>	commerce.
24	<i>Fivarotam-boky,</i>	librairie.
25	<i>Fivarotam-panafody,</i>	pharmacie.
26	<i>Fivarotan-tena,</i>	prostitution.
	♦ Dupli.,	Varobarotra

Figure 3 : Varotra (Rajaonarimanana 1995 : 325)⁸

⁸ Dans la typographie originale de NR, les sous-entrées ne sont pas séparées par des retours à la ligne et toute l'entrée lexicale apparaît en un seul bloc, formaté en drapeau sous la vedette.

On note, outre la richesse et la précision de l'information lexicale présentée, le haut degré d'organisation de la présentation elle-même⁹.

Particularités des différents dictionnaires

Abinal & Malzac (1888) (A&M)

Chez A&M, les différentes sous-entrées sont distinguées systématiquement par l'utilisation de caractères gras pour la vedette de chaque sous-entrée (qui apparaît, formatée en drapeau, à la suite de la vedette), tandis que les différents types de sous-entrées (formes fléchies, formes dérivées, composés, expressions) ne sont pas distinguées typographiquement entre elles¹⁰. Les exemples, par contre, sont distingués typographiquement : ils sont en italiques et regroupés dans le même paragraphe que la sous-entrée qu'ils illustrent. On notera aussi, chez A&M, que pour les formes préfixées en *Ci-* (où $C \in \{m, f, mp\}$), le préfixe de verbalisation *mi-* (réalisant la fonction lexicale P_0 dans le vocabulaire des fonctions lexicales (FL) de la théorie sens-texte (TST), cf. Mel'cuk, Clas et Polguère (1995)) est présenté en premier, suivi des préfixes de nominalisation d'action *fi-* (réalisant la FL S_0 de TST) et de nominalisation d'acteur *mpi-* (réalisant la FL S_1 de TST), alors que pour les formes relatives (circonfixées), c'est la forme nominalisée *fivarotana* (circonfixée en *fi--ana* et réalisant une fonction lexicale complexe $S_{0_{rel}}$) qui apparaît en premier et la forme verbalisée *ivarotana* (circonfixée en *i--ana* et réalisant une fonction lexicale complexe $P_{0_{rel}}$) qui apparaît en second. La forme *mpivarotana*, théoriquement possible (et qui pourrait signifier quelque chose comme «la personne chez qui, pour qui ou à cause de qui l'on vend») n'est jamais lexicalisée. Cet ordre de

⁹ Nous avons quelque peu modifié la typographie de façon à mieux séparer les différentes sous-entrées et nous avons numéroté les composés, phrasèmes et exemples pour faciliter les références croisées d'un dictionnaire à l'autre. Nous avons également modernisé l'orthographe d'Abinal et Malzac (1888). Tous les accents aigus ont ainsi été remplacés par des accents graves, conformément à l'usage contemporain. Cet accent note l'accent d'intensité, particulièrement marqué en malgache. Dans l'orthographe usuelle, on ne note l'accent que s'il ne suit pas les lois régulières d'accentuation. Ces lois sont fort simples : (i) pour les monosyllabes, accent sur cette syllabe; (ii) pour les formes polysyllabiques, accent sur la pénultième en général et sur l'antépénultième pour certains mots dont la dernière syllabe est *kV*, *trV* ou *nV* (< *k*, *t*, *r*, *n* fermant cette syllabe en proto-malgache). Nous avons tout de même conservé ici les accents qui ne seraient pas notés dans l'orthographe usuelle. Dans A&M, l'accent est marqué systématiquement sur le dernier mot de la vedette, et uniquement sur celui-ci, aussi bien pour l'entrée principale que pour les dérivés ou les exemples. Même dans le cas où deux exemples sont coordonnés métalinguistiquement, seul le dernier mot du deuxième exemple porte l'accent : dans «Varo-mahay mody et Varo-mièra», *miera* est accentué (alors qu'il ne le serait pas en orthographe usuelle), tandis que *mody* ne l'est pas (*et* est ici une conjonction métalinguistique). Dans NR, seule la vedette principale est accentuée. Dans RR, seuls sont accentués les mots qui portent l'accent en orthographe usuelle.

¹⁰ En fait, on remarque que, dans la typographie d'A&M, les sous-entrées perçues comme plus régulières (formes fléchies ou dérivées) ne sont pas séparées par un retour à la ligne et les formes dérivées dans ces sous-entrées ne sont pas en caractères gras, mais en petites majuscules, comme les renvois. Ainsi, au début de l'article *varotra*, les formes *FI—* et *MPI—*, qui sont des préfixes complexes de nominalisation régulièrement associés au préfixe complexe de verbalisation *MI—* (comme les préfixes complexes *fan(a)-* et *mpan(a)-* sont associés à *man(a)-*) sont groupées avec cette dernière dans un même paragraphe.

présentation (verbalisation avant nominalisation pour l'actif et nominalisation avant verbalisation pour le relatif ou circonstanciel) est assez systématique chez A&M et semble représenter une perception intuitive de l'importance ou de la fréquence des formes qui est conforme à nos observations.

Rajemisa-Raolison (1985) (RR)

Chez RR, les formes dérivées sont regroupées dans une rubrique spéciale de l'entrée. Elles apparaissent après le symbole «•,●» et l'abréviation *t.i.f.* (*teny iray fototra*, «mots de même racine»). Elles sont en caractères italiques gras et groupées en séries apparentées. Les séries sont séparées par des diagonales («/»). Les termes de ces séries ne sont pas explicités mais constituent autant de renvois aux entrées correspondantes¹¹. Les composés et les phrasèmes apparaissent en italiques, après le symbole «*» et l'abréviation *F.-p.* (*fombam-piteny*, «expressions» — littéralement «façons de dire») séparés par une diagonale «/». Les différents sens sont séparés par le symbole «♦». Les sens figurés sont indiqués par le symbole «◆» suivi de l'abréviation *h.o.* (*hevitra an'ohatra*, «sens figuré»)¹².

Notons enfin que RR est un dictionnaire illustré et qu'une entrée lexicale peut également faire référence à une ou plusieurs illustrations ou encore à un tableau¹³. Le lien entre entrée ou illustration est indiqué soit par une flèche pointant de l'entrée vers l'illustration, soit par une reprise, à l'intérieur de l'illustration, de la vedette de l'entrée. Nous n'avons pas tenu compte de cet aspect ici, puisqu'il s'agit de connaissances encyclopédiques plutôt que linguistiques, mais il est certain que dans une représentation multimédia, de telles représentations graphiques ont une grande importance. Une chose est certaine, comme l'institution RR est la seule à contenir de telles références, il n'y aurait aucun problème à coller celles-ci (ou, en fait, toute autre illustration pertinente) sous l'entrée lexicale appropriée¹⁴.

¹¹ Puisque, comme on l'a déjà mentionné, ce dictionnaire est le seul où les dérivés constituent des entrées autonomes apparaissant telles quelles à leur place alphabétique dans le dictionnaire.

¹² Les synonymes (il n'y en a pas dans cette entrée) apparaissent après le symbole «¶» et l'abréviation *M.h.* (*mitovy hevitra*, «sens équivalent»). Les proverbes (extrêmement nombreux et courants en malgache) apparaissent après le symbole «★» et l'abréviation *Ohab.* (*ohabolana*, «proverbe»).

¹³ Ainsi, pour *ohabolana* ('proverbe'), on a un tableau de proverbes classés par thèmes (RR, pp. 775-777), pour *varimbazaha* (litt. 'riz d'Européen', 'blé'), on a une photo de blé sur pied (RR, p. 1016).

¹⁴ Le lien entre information linguistique et information graphique est cependant loin d'être trivial dans RR, même pour un agent humain. Par exemple, à la même page (RR, p. 1016) on retrouve une photo de lémurien (un maki, *lemur catta*, *lemuroïde*) sans aucune entrée qui pointe vers elle, ni identification sur la photo. On doit donc parcourir l'ensemble des entrées de la page pour savoir laquelle est pertinente. Il s'agit de l'entrée «**Varika** a. : Karazana gidro kely, ny rambony misoratra mitapatapaka.» ('Espèce de petit lémurien dont la queue est striée.') Cette définition, qui correspond à la traduction qu'A&M donnent de *varika* («Petit Lémurien : le macoco à queue annelée, *Lemur catta*, L.») semblerait plutôt appropriée pour *maky* que pour *varika* (*lemur fulvus*). RR a peut-être repris A&M sans vérification. Il faut cependant noter que l'usage courant (surtout celui des citadins) ne respecte pas nécessairement celui des spécialistes.

Rajaonarimanana (1995) (NR)

Chez NR, l'organisation est plus simple que chez A&M ou RR. L'information lexicale est relativement moins riche : il y a beaucoup moins d'exemples, les dérivés ne sont pas mentionnés systématiquement et la forme de l'impératif (ici avec la consonne thématique *-t-*) doit être déduite de la forme du dérivé *fivarotana*. Par contre, on retrouve plus de termes associés aux réalités contemporaines (cf. les items 4, 9, 16, 17, 18) ainsi que des composés ou des phrasèmes eux-mêmes construits sur les dérivés (cf. les items 19 à 26). C'est cette récursivité qui rend les entrées, en apparence assez simples, difficiles à analyser aussi bien pour l'humain que pour un programme. En ce qui concerne l'organisation typographique, les composés et phrasèmes apparaissent en italiques minces, tandis que les dérivés apparaissent en italiques grasses, précédés du symbole «♦». Les dérivés de dérivés apparaissent en principe précédés du symbole «•» (mais cet usage n'est pas systématique).

Recouvrements des différents dictionnaires

On peut déjà constater à partir de cet exemple que la tradition lexicographique malgache ne le cède en rien à la tradition lexicographique française qui consiste à copier systématiquement ses prédécesseurs et ce, non seulement pour la nomenclature (où les données linguistiques elles-mêmes font qu'il est impossible d'échapper aux répétitions), mais pour les définitions et les exemples. Ainsi les items 2, 3, 6, 8, 9, 11 à 22 et 26 de RR reprennent essentiellement A&M, tandis que l'item 6, les items 3 et 12, les items 1, 7, 8, 10, 11 13 et 14 de NR reprennent respectivement A&M, RR et à la fois A&M et RR.

Sous des apparences superficiellement assez distinctes et des organisations diverses (syntaxiquement et typographiquement), on sent bien que ces dictionnaires disent à peu près la même chose sur un certain nombre de points et qu'ils se complètent sur d'autres. Ces différentes entrées contiennent essentiellement un noyau d'informations communes et des satellites spécifiques. C'est cette intuition que nous essayons de mettre en oeuvre. Ceci suggère la possibilité de poursuivre la tradition et de reprendre ces différentes données, en les fusionnant dans une base lexicale ouverte.

Syntaxe des entrées lexicales dans les dictionnaires traditionnels

Examinons de plus près la forme de ces entrées. Nous utiliserons des expressions régulières de façon à pouvoir en caractériser succinctement la forme générale¹⁵.

¹⁵ Ces expressions régulières ne caractérisent pas directement les entrées lexicales, mais en constituent une description abstraite et schématique en ce sens que les symboles qui y apparaissent sont aussi bien des catégories abstraites (par exemple *L, R, C, T*, etc.) que des symboles spécifiques (par exemple ♦, •, •, *t.i.f.*, •, etc.). De plus, nous ne tenons pas compte de certains détails de typographie, de ponctuation, etc. Les opérateurs propres employés dans ces expressions seront :

- l'opérateur unaire postfixé dit «étoile de Kleene» («*») qui signifie que le terme précédent peut-être répété un nombre indéfini de fois (y compris aucune);
- l'opérateur unaire postfixé («⁺») qui signifie que le terme précédent peut-être répété un nombre indéfini de fois (mais au moins une : $x^+ = xx^*$);

Abinal & Malzac (1888) (A&M)

L'entrée de *varotra*, de même que presque toutes les entrées dans A&M¹⁶ peut être décrite abstraitement au moyen de l'expression régulière suivante :

L(**R**)(**C**T*(**E**T)*(**F**C*)N*

(D(**C**)T*(**E**T)*(**F**C*)N*)*

(P(**C**)T*(**E**T)*(**N**))*

où

- L** est une vedette,
- R** est une racine¹⁷,
- C** une catégorie,
- T** une traduction,
- E** un exemple d'emploi,
- F** une forme fléchie (généralement l'impératif¹⁸),
- N** une annotation ou un renvoi,
- D** un affixe de dérivation ou une forme dérivée,
- P** une expression (ou un proverbe).

— l'opérateur n-aire de disjonction («{ $-_1 \mid -_2 \dots$ }») qui indique un choix exclusif entre les termes $-_1, -_2, \dots$ qu'il renferme;

— l'opérateur implicite de concaténation.

Rappelons que les parenthèses sont des opérateurs impropres qui ont deux rôles :

- 1°) indiquer l'optionnalité d'un terme et
- 2°) regrouper des termes en un seul.

¹⁶ Les entrées qui ne peuvent pas être décrites par ces expressions régulières sont des entrées exceptionnelles. Ou bien on y trouve une combinaison booléenne de termes, ou bien des termes apparaissant dans une position exceptionnelle. Par exemple «HALAVÏRINA, HALAVÏTINA, p. de Lavitra.» (p. 209, conjonction de vedettes), «LÈMBA (trivial), *adj.* Mort.» (p. 399, annotation immédiatement après la vedette) ou encore «KÒTAKA et KOTAKÒTAKA (*dupl.* plus usité), s. Tumulte, turbulence, révolution. Comme HOTAKOTAKA.» (p. 361, conjonction de vedettes et annotation immédiatement après l'entrée). Tenir compte de ces cas ne pose aucun problème de principe, mais aurait compliqué indûment l'expression.

¹⁷ Cette racine apparaît (entre parenthèses) quand la vedette n'est pas elle-même une forme racine, mais une forme dérivée idiosyncratique (par exemple «KITRANOTRÀNO (*trano*), s. Petites maisonnettes que font les enfants.» (p. 340). La dérivation en *ki-duplicatif* (ou *tsi-duplicatif*) pour indiquer un jeu impliquant le sens de la racine (ici *trano* : 'maison') est idiosyncratique) ou opaque (par exemple «MÀINA (*haina*), *adj.* Sec. desséché, tari; *fig.* désappointé, déchu, surpris, qui n'a rien à dire. [...]» (p. 419), «MÀINTY, (*inty*) *adj.* Noir. [...]» (ibidem), «TSATSÏAKA (*tsiaka*), s. Action de déchirer avec bruit, de faire du bruit en se déchirant, en se fendant. [...]» (p. 756). Le dérivé duplicatif attendu de *tsiaka* serait *tsiatsiaka*, qui existe effectivement et apparaît sous l'entrée de *tsiaka*).

¹⁸ L'impératif est une forme importante, puisqu'il est relativement imprédictible et que l'allomorphe du radical (incluant souvent une consonne finale) sur lequel il est formé sert de base dans les dérivations suffixales (et circonfixiales).

On remarque que cette description est divisée en trois zones :

(a) l'entrée principale :
 $L (R) (C T^* (E T)^* (F C)^*) N^*$,

(b) les dérivés :
 $(D (C) T^* (E T)^* (F C)^* N^*)^*$

et

(c) les expressions (ou proverbes) associés :
 $(P (C) T^* (E T)^* (N))^*$.

Ainsi sont bien distinguées par leur position respective les informations globales sur l'item lexical (sa catégorie, ses traductions, ses exemples d'emploi avec leurs traductions, ses formes flexionnelles, ses liens sémantiques à d'autres items, etc.), qui apparaissent dans la première zone et les informations morphologiques ou phraséologiques sur ses dérivés ou ses expressions associées qui apparaissent respectivement dans sa deuxième et sa troisième zone¹⁹.

Rajemisa-Raolison (1985) (RR)

Les entrées de RR peuvent être décrites par l'expression régulière suivante :

$L (d) (R) (C) (S (: E^+)) (\diamond S (: E^+))^*$
 $(\bullet, \bullet \text{ t.i.f. } D^+ (/ D^*)^*)$
 $(\bullet \text{ F.p. } : P (: S (: E^+) \diamond H.o. S) (/ P (: S (: E^+) \diamond H.o. S))^*)^*$
 $(\star \text{ Ohab. } (O : S)^+)$
 $(\parallel \text{ M.h. } : M^+)$

où

L est une vedette,
d est une indication dialectale ou étymologique²⁰,
R est une racine,
C une catégorie,
S une définition,
E un exemple d'emploi,
D une forme dérivée,
P une expression,
M une forme synonyme,
♦ le délimiteur des significations différentes,

¹⁹ Ces informations, de même que les liens facultatifs vers la racine ou vers des synonymes globaux de la première zone, constitueraient autant d'hyperliens vers d'autres formes dans une représentation en hypertexte.

²⁰ Par exemple : Marotoa (*fr. marteau*) a. : Tantanam-by kely. ('petit marteau de fer').

- ,• *t.i.f.* le délimiteur de la liste des dérivés (*teny iray fototra*, «mots de même racine»),
- *F.p.* le délimiteur de la liste des expressions (*fombam-piteny*, «manières de parler»),
- ★ *Ohab.* le délimiteur de la liste des proverbes (*ohabolana*)

et

- ¶ *M.h.* le délimiteur de la liste des synonymes²¹ (*mitovy hevitra*, «même sens»).

Chacun des termes complexes de l'expression régulière correspond donc à une zone définie :

- (a) l'entrée principale :
 $L(d)(R)(C)(S(:E^+))(\blacklozenge S(:E^+))^*$,
- (b) les dérivés :
 $(\bullet, \bullet \text{ t.i.f. } D^+ (I D^*)^*)$,
- (c) les expressions associées :
 $(\bullet F.p. : P(:S(:E^+) \blacklozenge HaS)(P(:S(:E^+) \blacklozenge HaS)^*))$,
- (d) les proverbes associés :
 $(\star Ohab. (O : S)^+)$,
- (e) les synonymes (et antonymes) :
 $(\¶ M.h. : M^+)$

On remarque que les zones sont plus nombreuses et plus faciles à identifier, du fait qu'elles sont systématiquement indiquées au niveau typographique.

²¹ Les antonymes (*mifanohi-kevitra*), peu nombreux, sont aussi notés sous cette rubrique. On remarque qu'il n'y a pas de lien entre cette liste et la liste des définitions dans l'entrée principale. Généralement, cette liste fait référence au sens premier du terme (i.e. le premier élément S de l'entrée principale). Comme les homonymes font l'objet d'entrées distinctes, ceci ne pose pas de problème (du moins pas à un humain).

Rajaonarimanana (1995) (NR)

Quant aux entrées dans NR, elles peuvent être décrites par l'expression régulière suivante :

$$L \quad (R) (C) (T) ((n)(C) (n)(T) (F) (C) (T)) \{ (E T) | N \}^* \\ (\diamond \{ D | \text{Dupli.} (D) \} (C) (T) ((n)(C) (n)(T) (F) (C) (T)) \\ \{ (E T) | N \}^* | (\bullet D) (C) (T) ((n)(C) (n)(T) (F) (C) (T)) \{ (E T) | N \}^* \}^*)^*$$

où

- L** est une vedette,
- R** est une racine,
- C** une catégorie,
- T** une traduction,
- n** un numéro de sous-entrée,
- F** une forme fléchie (généralement l'impératif),
- E** un exemple d'emploi ou une expression associée,
- D** une forme dérivée,
- Dupli.** l'indication d'une forme redoublée,
- N** une annotation indiquant un synonyme, un antonyme, une forme dialectale, une variante, un renvoi, etc.,
- ♦** le délimiteur des dérivés,
- le délimiteur des sous-dérivés.

Cette expression est plus complexe que les précédentes du fait que l'entrée n'est pas organisée linéairement en zones distinctes, mais hiérarchiquement en sous-entrées²². Cette récursivité (que des dictionnaires rigoureux comme le DEC (Mel'cuk et al. 1984, 1988, 1992; Mel'cuk, Clas et Polguère 1995) excluent par principe) pose des problèmes de lecture et d'interprétation assez difficiles. De plus, les exemples d'emploi ne sont pas distingués typographiquement des expressions.

²² Il serait donc plus simple de représenter les entrées de NR au moyen d'un réseau de transitions récursif ou d'une grammaire hors-contexte, par exemple :

ENTRÉE → Lex (Rac) Cat Trad (FORME) (ILLUSTRATION) (DÉRIVÉS) (SOUS-ENTRÉES)

FORME → Lex (Cat) (Trad)

ILLUSTRATION → (Ex Trad) (Note) (ILLUSTRATION)

DÉRIVÉS → DÉRIVÉ (SOUS-DÉRIVÉS) (DÉRIVÉS)

DÉRIVÉ → ♦ SOUS-ENTRÉE

DÉRIVÉ → ♦ Duplicatif (SOUS-ENTRÉE)

SOUS-DÉRIVÉS → • DÉRIVÉ (SOUS-DÉRIVÉS)

SOUS-ENTRÉES → SOUS-ENTRÉE (SOUS-ENTRÉES)

SOUS-ENTRÉE → (n) Lex (Cat) (Trad) (FORME) (ILLUSTRATION)

où Lex est une forme quelconque, Rac est une racine, Trad est une traduction, Ex est un exemple, Note est une annotation quelconque, Duplicatif est une indication de duplicatif, n est un nombre.

Fusion et filtrage

Si l'on disposait de ces dictionnaires sous format électronique (ce qui n'est pas le cas, notons-le), on pourrait, étant donné une telle caractérisation, les fusionner en une base lexicale complexe gardant (ou non) la trace de la source de ses données²³.

On se contentera d'utiliser ici la technique classique et heuristiquement productive du *fantasme*. C'est-à-dire que l'on présupposera tout simplement que l'on dispose de ces dictionnaires sous format électronique. En pratique, le problème de saisie se combine à celui du *filtrage*. En fait, grâce aux outils que nous avons développés, les deux opérations peuvent être combinées jusqu'à un certain point. Plutôt que de saisir globalement chacun des dictionnaires, pour les filtrer ensuite, on en a saisi des fragments suffisants pour construire semi-automatiquement un squelette de base de données lexicales contenant plus de 10 000 entrées de base (racines ou dérivés). Ce squelette a lui-même été retravaillé de diverses manières et a donné lieu à plusieurs expérimentations. Il est actuellement enrichi semi-automatiquement de façon systématique à partir de données contenues dans les dictionnaires traditionnels (cf. outre A&M, NR et RR, Malzac 1908; Profita 1969; Rakotonaivo 1996), et dans d'autres institutions lexicales, comme les lexiques-grammaires (Rabenilaina 1987, 1996). Chaque entrée lexicale de chacun des dictionnaires peut être analysée selon la grammaire propre au dictionnaire, puis standardisée selon un schéma commun. Les différents schémas ainsi obtenus peuvent alors être fusionnés en un schéma unique²⁴. C'est précisément à la caractérisation de tels schémas que serviront les institutions lexicales.

Nous avons développé, à titre expérimental, quelques maquettes de programmes d'analyse et de fusion pour les entrées lexicales dans ces différents formats. On trouvera, en annexe, le résultat de l'application d'un de ces programmes aux trois entrées de *varotra*.

²³ La saisie de ces textes pose toutes sortes de problèmes techniques. Dans le cas d'A&M, par exemple, l'impression la plus récente (1993) reproduit photostatiquement l'édition de 1888. Une saisie par lecteur optique est donc pratiquement exclue dans ce cas, du fait de la faible qualité graphique de l'original. Par ailleurs, il y a les délicats problèmes de droits d'auteur posés par les dictionnaires récents. Certains considèrent que même leur nomenclature devrait être protégée, ce qui est soit vide de sens (ajouter ou enlever un seul terme ferait que l'on aurait une autre nomenclature, ou alors il faudrait établir des mesures de distance entre nomenclatures, en précisant des seuils au-delà desquels des nomenclatures seraient distinctes), soit ridicule (tout nouveau traitement d'un terme déjà couvert par un dictionnaire serait exclu). Comme la nomenclature, les traductions et même les exemples sont souvent eux-mêmes empruntés à A&M, ces problèmes sont loin d'être évidents. Le fait de noter la source des données permettrait probablement de contourner certains problèmes de droits d'auteur, à condition de ne pas en faire un usage commercial, bien sûr.

²⁴ On pourrait penser que les techniques classiques d'alignement (Gale et Church 1993; Blank 1995) seraient utiles pour appairer les entrées correspondantes des différents dictionnaires. Tel n'est pas le cas. En effet, l'équivalence entre dictionnaires n'est pas du tout du même type que l'équivalence entre un texte et sa traduction. Un dictionnaire n'est pas la «traduction» d'un autre, même si les emprunts, voire les traductions (cf. RR, qui utilise parfois comme définition d'un terme malgache, la retraduction en malgache de l'équivalent français du terme donné par A&M) sont fréquents. Il n'y a pas de correspondance statistiquement significative entre dictionnaires. L'appariement, s'il est possible, doit se faire de façon symbolique.

3. INSTITUTIONS LEXICALES

Une *institution* est l'ensemble des prédicats implicites ou explicites (mais le plus souvent implicites) qui caractérisent un cadre théorique ou descriptif²⁵. Appliquée au domaine lexical, cette notion suggère l'idée suivante. Si l'on sait de quoi une institution lexicale (i.e. un type de lexique, de dictionnaire, etc.) parle et si l'on sait traduire d'une institution à l'autre, on peut fusionner des informations provenant de plusieurs institutions lexicales²⁶. Autrement dit, si la sémantique est claire, la syntaxe (comme l'intendance) devrait pouvoir suivre. Ainsi, on sent bien que, sous les divergences de forme, les trois dictionnaires traditionnels que nous avons mentionnés parlent bien (du moins partiellement) des mêmes choses. Si l'on arrive à déterminer de façon plus précise quel est le domaine couvert par chacun de ces types de description, on pourra déterminer les conditions permettant de les fusionner.

Il y a deux bonnes nouvelles ici. La première, qui peut sembler paradoxale, c'est que si deux institutions ne parlent pas de la même chose (i.e. si leurs domaines sont entièrement distincts, s'il n'y a pas d'interface entre elles), on peut toujours les combiner. Par exemple, si chacun des dictionnaires traditionnels contenait des informations entièrement distinctes, on pourrait tous les combiner facilement. Évidemment, pour qu'une combinaison soit intéressante, il faut qu'il y ait une interface²⁷. Cependant cette interface n'a pas à être *totale*. Les parties d'une institution dont le domaine n'a pas d'équivalent dans une autre institution sont combinables à cette dernière sans problème. La deuxième, c'est que, lorsque deux institutions parlent effectivement de la même chose, on peut soit fusionner les informations, lorsqu'on sait qu'elles sont compatibles et qu'on sait les fusionner, soit tout simplement les cumuler, en en indiquant la source²⁸ quand on ne sait pas si elles sont compatibles ou comment les fusionner.

Ainsi, si l'on sait

- (a) que l'institution lexicale A&M attribue à une entrée dont la VEDETTE est **VÀROTRA** la CATÉGORIE *s.* et la TRADUCTION «Vente, commerce, négoce, trafic, marchandises.»,

²⁵ Cf. Blache & Morin (1997).

²⁶ L'indéfini (*on, l'on*) dans ces énoncés est voulu. Les agents peuvent être humains, machines, ou les deux. En fait, ce que la philosophie des institutions lexicales favorise, ce sont justement des traitements interactifs, coopératifs, partiellement automatisés. Ceci permet une très grande souplesse (et une indépendance totale) par rapport aux grandes entreprises de normalisation (HTML, SGML, XML, etc.).

²⁷ On peut imaginer des cas de figure où un tel assemblage hétéroclite aurait de l'intérêt, une encyclopédie, par exemple, où les articles de littérature comparée et ceux de géologie auraient peu de chance de porter sur des domaines en intersection (encore que les formes — strates, couches, etc. — puissent être en intersection). Mais une véritable encyclopédie ne saurait se limiter à ces deux domaines et il est probable, que, de proche en proche, les domaines soient finalement connexes.

²⁸ Cette indication sert, en quelque sorte, de ceinture de protection contre l'incohérence. Associer à une entité E deux propriétés P et P' qui sont incompatibles produit une entité incohérente. Mais associer à E deux propriétés X(P) et Y(P') peut permettre d'éviter cela jusqu'à ce qu'un agent plus compétent parvienne à les concilier.

- (b) que l'institution lexicale NR attribue à une entrée dont la VEDETTE est **vàrotra** la CATÉGORIE n. et la TRADUCTION «Vente, commerce, marchandises.»,
- (c) que l'institution lexicale RR attribue à une entrée dont la VEDETTE est **Varotra** la CATÉGORIE a. et les DÉFINITIONS «Fanakalozan-javatra amin'ny vola, hitadiavan-tombony» et «Entana amidy»

et que l'on sait

- (d) identifier les vedettes **VÀROTRA** de A&M, **vàrotra** de NR et **Varotra** de RR et les fusionner en une vedette 'varotra',
- (e) identifier les catégories s. de A&M, n. de NR et a. de RR et les fusionner en une catégorie 'N',

mais que l'on ne sait pas (pour le moment²⁹)

- (f) fusionner «Vente, commerce, négoce, trafic, marchandises.» de A&M et «Vente, commerce, marchandises.» de NR (ni même si elles sont compatibles)

ni

- (g) si les TRADUCTIONS de A&M et NR sont compatibles avec les DÉFINITIONS de RR,

on pourra déjà créer, dans une institution lexicale A&MNRRR, une entrée lexicale partielle comme la suivante³⁰ :

²⁹ Cette qualification est importante. Les institutions lexicales sont des entités souples et dynamiques. Si, par la suite, un agent (humain ou non) qui sait fusionner ces définitions intervient, il pourra les modifier. Par exemple, un lexicographe ou un programme maximaliste, qui ferait l'union des termes de deux ensembles de traductions pourrait fusionner ces deux ensembles en un seul, équivalent au premier (puisque le deuxième est un sous-ensemble du premier). Un agent qui pourrait faire correspondre une définition en langue naturelle et une définition de type DEC pourrait produire la définition et les fonctions lexicales suivantes. La définition est donnée en pseudo-malgache métalinguistique, suivie d'un équivalent en pseudo-français métalinguistique, les indices indiquent des étiquettes sémantiques qui constituent autant de contraintes de type (cf. Miličević 1997).

Définition

fiasa an'ny fiaraha-monina : ~ Y entana amin' Z olona amin' W vola X olona
 (action sociale : X personne ~ Y marchandise à Z personne pour W argent)
 Fonctions lexicales : S₀(varotra) = varotra, S₂(varotra) = varotra

³⁰ La forme spécifique de la représentation n'a pas d'importance à ce niveau. En fait, les objets créés dans nos maquettes sont des objets prolog, que l'on pourrait aisément traduire en objets HTML, SGML ou XML au besoin. On se rend compte qu'une bonne partie du travail créateur de traitement de l'information (qu'elle soit lexicale ou non) consiste à manipuler et à expérimenter avec des représentations diverses. Le fait d'utiliser des représentations prolog pour les institutions lexicales facilite ce travail. Les institutions lexicales sont des entités plus abstraites et plus souples (et surtout moins lourdes) que les DTD de SGML, par exemple, même s'il y a évidemment des liens entre les deux. Ce ne sont pas des normes, mais des caractérisations abstraites des domaines sémantiques recouverts par des objets descriptifs. Elles

VEDETTE	varotra	
CATÉGORIE	N	
TRADUCTIONS	"Vente, commerce, négoce, trafic, marchandises."	(A&M)
	"Vente, commerce, marchandises.",	(NR)
DÉFINITIONS	"Fanakalozan-javatra amin'ny vola, hitadiavan-tombony"	(RR)
	"Entana amidy"	(RR)

Figure 4 : Entrée lexicale partiellement fusionnée pour *varotra*

Quels sont donc les domaines recouverts par ces trois dictionnaires ? Quelle est la structure conceptuelle des entrées lexicales qu'on y retrouve ? Une entrée lexicale complète³¹ se compose obligatoirement d'une VEDETTE, d'une CATÉGORIE et d'une liste de TRADUCTIONS ou de DÉFINITIONS³². Toute entrée lexicale se construira donc autour d'un squelette [VEDETTE, CATÉGORIE, TRADUCTIONS ou DÉFINITIONS]. Les éléments de ce squelette forment respectivement le noyau des trois zones [PHONOLOGIQUE/GRAPHIQUE, SYNTAXIQUE, SÉMANTIQUE].

Évidemment, un tel squelette est très incomplet, particulièrement aux niveaux SYNTAXIQUE et SÉMANTIQUE. Tout ce qu'il contient au niveau SYNTAXIQUE, c'est une *partie du discours*. Bien qu'une partie du discours constitue un type syntaxique pauvre, ce caractère peu contraignant lui permet de se combiner librement avec de l'information syntaxique plus riche (liste et formes d'arguments) provenant d'une autre source³³. Quant à la SÉMANTIQUE, une traduction ou une définition n'est évidemment pas une représentation satisfaisante. Mais, suivant en cela notre tendance au retardement des tâches difficiles (équivalent en quelque sorte à de l'évaluation paresseuse), on considère que les traductions ou les définitions ne sont que des *noms* que nous attribuons à la représentation sémantique

sont orientées vers des traitements spécifiques interactifs où chacun des agents (humain ou machine) est chargé de tâches pour lesquelles il est mieux adapté (les tâches pouvant être redistribuées dynamiquement) et non vers des traitements de masse entièrement automatisés.

³¹ Une entrée lexicale incomplète est une entrée où l'une ou l'autre de ces rubriques manque. Dans le cas des entrées lexicales principales, la VEDETTE est toujours présente, mais pour les sous-entrées (qui constituent les DÉRIVÉS), elle peut être absente ou suggérée par des affixes (*mi-*, *fi-*, *mpi-*, etc.) ou des opérations (*Duplicatif*). Les autres cas d'entrées lexicales incomplètes sont des cas où la CATÉGORIE et/ou les TRADUCTIONS/DÉFINITIONS sont absentes. Dans les entrées principales, elles sont remplacées par un simple RENVOI (un type particulier d'ANNOTATION) à une entrée lexicale complète. Dans les sous-entrées, elles sont implicites (et généralement reconstructibles à partir des affixes ou des opérations).

³² TRADUCTIONS et DÉFINITIONS sont des composantes conceptuellement équivalentes, puisqu'on peut voir une définition comme une traduction dans un métalangage de définition qui peut utiliser soit le vocabulaire et la syntaxe de la langue décrite elle-même, soit celui d'une autre langue. En fait, elles correspondent l'une ou l'autre à la zone sémantique de l'entrée.

³³ Par exemple, si l'on sait qu'un lexique-grammaire (Rabenilaina 1987, 1996) contient de l'information sur la distribution syntaxique des arguments d'un prédicat et que les dictionnaires traditionnels ne contiennent aucune information sur ce domaine, on peut tout simplement coller l'information distributionnelle provenant du lexique-grammaire et l'information provenant des dictionnaires traditionnels. C'est ce que nous examinons dans un autre travail (Morin 1997).

effective. Que celle-ci existe ou pas n'invalide pas l'usage de ces chaînes comme des noms, ni l'usage de ceux-ci pour distinguer des acceptions.

Une entrée lexicale ne contenant que ce squelette est une *entrée lexicale simple*. Une entrée lexicale peut aussi être composite et comporter un ensemble de rubriques facultatives. Elle peut contenir une liste de DÉRIVÉS, une liste d'EXPRESSIONS, une liste d'EXEMPLES et/ou une liste d'ANNOTATIONS. On ajoutera donc implicitement au squelette ci-haut de telles rubriques de façon à obtenir pour chaque entrée lexicale, une structure virtuelle comme la suivante :

VEDETTE	<i>vedette</i>
CATÉGORIE	<i>catégorie</i>
TRADUCTIONS/DÉFINITIONS	<i>liste de traductions/définitions</i>
DÉRIVÉS	<i>liste de dérivés</i>
EXPRESSIONS	<i>liste d'expressions</i>
EXEMPLES	<i>liste d'exemples</i>
ANNOTATIONS	<i>liste d'annotations</i>

Figure 5 : Squelette d'une entrée lexicale fusionnée

EXPRESSIONS et EXEMPLES

Les entrées des dictionnaires traditionnels peuvent contenir des listes d'EXPRESSIONS associées et des listes d'EXEMPLES d'emploi. Les EXPRESSIONS et EXEMPLES sont distingués typographiquement des DÉRIVÉS dans chacun des trois dictionnaires. Chaque élément de ces EXPRESSIONS ou EXEMPLES est formé d'une VEDETTE, elle même associée à une liste de TRADUCTIONS/DÉFINITIONS et/ou à une liste d'ANNOTATIONS³⁴. Les EXPRESSIONS sont distinguées typographiquement des EXEMPLES

³⁴ Pour les EXPRESSIONS, on a généralement aussi une CATÉGORIE chez A&M. On notera aussi que les PROVERBES sont traités comme des EXPRESSIONS par A&M, alors que RR les isole en fin d'entrée (avant la liste des synonymes) et leur donne un statut particulier. Comme nous ne nous intéressons ici qu'à la dimension strictement linguistique, nous n'avons pas créé de rubrique particulière pour les proverbes qui, du fait de leur structure particulière, seront classés avec les exemples d'emploi.

aussi bien dans A&M³⁵ que dans RR³⁶, mais pas dans NR³⁷. Les EXPRESSIONS peuvent être formées d'un seul mot-forme, ou d'une séquence de mots-formes et être plus ou moins compositionnelles.

Expressions compositionnelles, semi-compositionnelles et non compositionnelles

À propos de la distinction entre *expressions compositionnelles* (dont le sens {est représentable, peut être reconstruit}) à partir de celui des composantes et de règles générales, *expressions non compositionnelles* (dont le sens \b\bc\{\(\a\ac(n'est pas représentable, ne peut pas être reconstruit)) à partir de celui des composantes et de règles générales) et *expressions semi-compositionnelles* (dont le sens \b\bc\{\(\a\ac(n'est représentable, ne peut être reconstruit)) que partiellement à partir de celui des composantes et de règles générales), il y a plusieurs choses à dire.

Premièrement, elle est souvent assez difficile à tracer, même dans des cas simples. Par exemple, si *mpivaro-boky* (*mpi-varotra-boky*, 'AGENT-commerce-livre' = 'libraire') est compositionnel, qu'en est-il de *mpivarotena* (*mpi-varotra-tena*, 'AGENT-commerce-{personne, corps, substance, éléments d'une chose, PRONOM RÉFLÉCHI}' = 'prostitué(e)') ? On voudrait certainement dire que ce dernier est «plus» compositionnel que *mitena* (*mi-tena*, 'VERBEPRÉSENT-{personne, corps, substance, éléments d'une chose, PRONOM RÉFLÉCHI}' = 'porter tous les jours les mêmes habits'), qui ne serait pourtant pas nécessairement totalement non compositionnel. Il y aurait donc des degrés de compositionnalité à établir.

Deuxièmement, si l'on savait construire, à partir des définitions présentes dans les institutions sources, une définition dans un langage standard, les (sous-)entrées correspondantes ne contiendraient que de l'information *héritée* pour les expressions *compositionnelles*, que de l'information *spécifiée* pour les expressions *non compositionnelles* et un mélange des deux pour les expressions *semi-compositionnelles*. Dès lors, les degrés de compositionnalité correspondraient au rapport entre (nombre de)

³⁵ Où la **vedette** de chaque expression est en caractères gras et débute un nouveau paragraphe, alors que les exemples sont en minces et viennent immédiatement à la suite de la forme qu'ils illustrent.

La forme illustrée par l'exemple y est marquée en italiques. Ainsi dans l'entrée de *varotra* ci-haut, la forme **ivarotana** est illustrée par un exemple :

«**Ivarotana**, *rel.* Ariary telopolo no *nivarotako* ny omby mifahiko. J'ai vendu trente piastres le bœuf que j'avais engraisé.»

où *nivarotako* (= n_{PASSÉ}+ivarotana+ah_{PRONOM}, 1 P.S.) est la forme du passé (préfixe *n-*), avec AGENT enclitique de première personne du singulier d'*ivarotana*.

³⁶ Où les expressions sont regroupées, après le délimiteur «• *F.p.*» et sont notées, avec leur *vedette* en italiques, immédiatement après le séparateur «/», et avant une définition et/ou des annotations en caractères minces, alors que les *exemples* sont entièrement en italiques.

³⁷ Où *expressions* et *exemples* d'emploi sont tous en italiques et non démarqués les uns des autres, avec les traductions ou annotations pertinentes en caractères minces. En fait, comme les EXEMPLES sont très peu nombreux dans ce dernier dictionnaire, on peut se contenter d'une procédure très simple pour les isoler, quitte à reclassifier certaines formes par la suite. Dans un premier temps, la procédure que nous utilisons, qui est efficace quel que soit l'agent qui l'effectue (humain ou non), consiste à considérer comme un EXEMPLE uniquement les entrées dont la VEDETTE a la forme extérieure d'une phrase, c'est-à-dire se termine par une ponctuation forte, ou contient une ponctuation quelconque.

composantes héritées et (nombre de) composantes spécifiées. Cependant, cette opération est, en général, très difficile même pour les agents humains (et *a fortiori* pour les machines).

Troisièmement, dans une base lexicale à accès multiples, cette distinction perd beaucoup de son importance. En effet, contrairement à la compositionnalité syntaxique, où une expression syntaxique compositionnelle est aussi bien analysable que synthétisable et n'a pas à être apprise individuellement, dans la compositionnalité lexicale en général, une expression lexicale compositionnelle analysable ou synthétisable de façon univoque doit être apprise. Ainsi, même si *mpivarokena* (*mpi-varotra-hena*, 'AGENT-commerce-{viande, chair}' = 'boucher') est une expression compositionnelle, le locuteur malgache doit apprendre qu'elle a bien ce sens et que c'est celle-là et pas une autre qui a ce sens particulier³⁸. Donc, compositionnel ou non, tout sens lexical doit être appris et stocké quelque part. La seule différence éventuelle, quand on sera en mesure de représenter systématiquement le sens lexical, sera que le sens des expressions compositionnelles sera représentable de façon beaucoup plus compacte que le sens des expressions non compositionnelles, sera plus facile à apprendre, à reconstruire en cas de perte, etc.

DÉRIVÉS

Une entrée lexicale peut aussi contenir une liste de DÉRIVÉS. On a vu que si celle-ci n'est qu'une simple liste structurée de formes, qui constitue autant de renvois, chez RR, elle peut être une liste de sous-entrées chez A&M. Elle peut même être récursive chez NR et contenir des dérivés de dérivés. Dans une base lexicale électronique (comme dans un dictionnaire formalisé du type DEC), cette récursivité sera éliminée, puisqu'il n'y aura pas de sous-entrées. On ne conservera que des *liens*, comme chez RR. Cependant, lors de la création de la base, il est plus pratique de suivre pas à pas les entrées lexicales des institutions sources et de créer des entrées complexes, qui pourront être «aplaties» par la suite. Un DÉRIVÉ est donc tout simplement une (sous-)entrée³⁹ apte à se détacher de son entrée d'origine en acquérant des liens bidirectionnels⁴⁰.

Les DÉRIVÉS se divisent en dérivés par *affixation* et dérivés par *réduplication*. Les deux types sont bien distingués aussi bien chez A&M, que chez NR⁴¹. Ce qui varie de

³⁸ De même, en français, si *fruitier* 'marchand de fruits' est compositionnel (par opposition à *boucher*, qui est non compositionnel), il faut apprendre que cette expression existe effectivement avec ce sens, alors que *légumier*, qui serait tout aussi compositionnel, n'a le sens de 'marchand de légumes' qu'en Belgique, à notre connaissance.

³⁹ Chez A&M, les sous-entrées des dérivés ne contiennent généralement pas de composés, ni de phrasèmes associés. Mais ils peuvent contenir des exemples. Chez Rajaonarimanana, les dérivés forment des sous-entrées de plein droit, comme on peut le voir ci-dessus pour *mpivarotra* et *fivarotana*.

⁴⁰ Du dérivé à la forme source (généralement une racine, mais éventuellement un dérivé) et de la forme source au dérivé.

⁴¹ Chez RR, comme la liste des dérivés est une simple liste de formes complètes, on doit disposer d'un analyseur morphologique pour identifier le type morphologique de chacune des formes et sa décomposition. Comme la dérivation est assez complexe, mais régulière, en malgache, nous avons pu construire un tel analyseur que nous comptons appliquer aux listes de formes dérivées de RR (saisies manuellement). En fait donc, il devrait être possible de reconstruire la forme complète d'un dérivé à partir de sa source et de son type, tout comme de reconstruire son type à partir de sa forme complète et de celle de sa source. Quant à la sémantique

dictionnaire à dictionnaire et même d'entrée à entrée dans un même dictionnaire, c'est la façon dont les formes sont présentées. Pour les formes affixées, on donne parfois simplement l'affixe, parfois une forme complète (cf. **Mi—**, **Maha—** vs **Fivaròtana**, **Ivaròtana** dans l'entrée *varotra* d'A&M ci-haut ou **Mi-** vs **Mpivarotra** et **Fivarotana** dans l'entrée *varotra* de NR⁴²). Pour les formes redoublées, elles sont toujours associées à une annotation (qui peut varier de forme) dans A&M. Dans NR, elles apparaissent parfois à la suite d'une annotation (en principe, l'abréviation «Dupli.»), parfois, elles restent implicites, seule l'annotation étant donnée.

On pourra donc ajouter au squelette d'entrée lexicale ci-haut une zone de DÉRIVÉS. Chaque DÉRIVÉ aura un TYPE MORPHOLOGIQUE (*préfixe*, *suffixe*, *circonfixe* ou *infixe* pour les dérivés affixaux et *duplicatif* pour les formes redoublées), en plus des autres rubriques communes à toutes les entrées : une VEDETTE (la forme dérivée complète), une CATÉGORIE⁴³, une liste (éventuellement vide) de TRADUCTIONS/DÉFINITIONS. Pour les dérivés affixaux, la forme de l'affixe pourra apparaître comme valeur de l'attribut correspondant au TYPE (*préfixe*, *suffixe*, *circonfixe* ou *infixe*).

Si donc on sait traduire les dérivés de chacune des institutions lexicales A&M, RR et NR, on peut les fusionner dans une institution A&MRRNR⁴⁴. Quand on détachera les dérivés de l'entrée principale où ils ont été créés, on les remplacera par un ensemble de liens correspondant aux différentes fonctions morphologiques possibles (i.e. les différents types de dérivés), mais également aux fonctions sémantiques (\approx les *fonctions lexicales* de la lexicographie explicative et combinatoire, Mel'cuk, Clas et Polguère (1995)) que ces différents types de dérivés instancient. Chaque dérivé sera donc accessible, depuis une racine par plusieurs chemins. Inversement, dans l'entrée du dérivé, les fonctions inverses (morphologiques ou sémantiques) retourneront l'entrée lexicale de la racine. Par exemple, appliquées à l'entrée *varotra*, la fonction morphologique *mp-* (ou *préfixe_{mp-}*), de même que la fonction sémantique *nominalisation-S_I* retourneront toutes deux le dérivé *mpivarotra* avec son entrée lexicale. Et la fonction *mp⁻¹* (ou *préfixe_{mp⁻¹}*) de même que la fonction *S_I⁻¹* appliquées à l'entrée *mpivarotra* retourneront l'entrée *varotra*. Ces fonctions inverses, particulièrement utiles pour l'apprenant de la langue, aussi bien que pour les programmes d'analyse, sont créées automatiquement. Étant donnée une racine *r*, on peut systématiquement créer un ensemble de *fonctions morphologiques* et de *fonctions sémantiques* régulièrement associées.

associée, elle devrait également être restructurable, du moins pour les affixes correspondant à des *fonctions lexicales* (au sens de la théorie sens-texte) précises, ce qui est le cas de la plupart des affixes en malgache (cf. infra).

⁴² Pour un même affixe, *mi-*, par exemple, on trouve dans RR 89 % d'entrées (957) avec l'affixe comme vedette du dérivé et 11 % (117) avec la forme complète. Comme cet affixe à finale vocalique n'entraîne pas d'allomorphie (sauf devant une base à «i» initial où les deux «i» se fusionnent généralement, mais pas toujours : *mi + iray = miray*, mais *mi + isa = miisa*), les formes complètes sont tout à fait prédictibles.

⁴³ Dans le cas où une institution source ne mentionne pas la catégorie du dérivé, celle-ci est toujours facile à reconstruire à partir du TYPE du DÉRIVÉ et de la CATÉGORIE de l'entrée contenant ce DÉRIVÉ. Pour les dérivés affixaux, chaque affixe définit la CATÉGORIE de son résultat. Pour les formes redoublées, le résultat est de la même CATÉGORIE que la forme de départ.

⁴⁴ Voir la zone des DÉRIVÉS de l'annexe qui donne une représentation externe du résultat de la fusion des trois entrées de *varotra*.

FONCTION MORPHOLOGIQUE	TYPE	FONCTION SÉMANTIQUE
<i>préfixe</i> _{an-}	RADICAL → V _{an}	<i>actif</i>
<i>préfixe</i> _{i-}	RADICAL → V _i	<i>moyen</i>
<i>préfixe</i> _{a-}	RADICAL → V _a	<i>statif</i>
<i>suffixe</i> _{-ina}	RADICAL _{thème} → V _{ina}	<i>passif</i>
<i>suffixe</i> _{-ana}	RADICAL _{thème} → V _{ana}	<i>passif</i>
<i>circonfixe</i> _{an--ana}	RADICAL _{thème} → V _{an-ana}	<i>circonstanciel</i>
<i>circonfixe</i> _{i--ana}	RADICAL _{thème} → V _{i-ana}	<i>circonstanciel</i>
<i>préfixe</i> _{tafa-}	RADICAL → X _{tafa}	<i>participe passif perfectif</i>
<i>préfixe</i> _{voa-}	RADICAL → X _{voa}	<i>participe passif perfectif</i>
<i>préfixe</i> _{f-}	V _{an/i} → N _f	<i>S0</i>
<i>préfixe</i> _{h-}	V _a → N _h	<i>S0</i>
<i>préfixe</i> _{mp-}	V* _{ina/ana} → N _{mp}	<i>S1</i>
<i>préfixe</i> _{ah-}	V _a → V _{ah}	<i>abilitatif</i>
<i>préfixe</i> _{amp-}	V* _{ina/ana} → V _{amp}	<i>causatif</i>
<i>préfixe</i> _{ank-}	V _{i/a} → V _{ank}	<i>transformatif</i>
<i>préfixe</i> _{if-}	V* _i → V _{if}	<i>réciproque</i>
<i>préfixe</i> _{ih-}	V _a → V _{ih}	<i>progressif</i>
<i>préfixe</i> _{m-}	V* _{ina/ana} → V _m	<i>présent</i>
<i>préfixe</i> _{n-}	V → V _n	<i>passé</i>
<i>préfixe</i> _{t-}	Prép → Prép	<i>passé</i>
<i>préfixe</i> _{h-}	V → V _h	<i>futur</i>
<i>redoublement</i>	RADICAL → RADICAL _{dupl}	<i>diminutif, fréquentatif, intensif, etc.</i>
<i>préfixe</i> _{ki-}	RADICAL _{dupl} → X _{ki}	<i>diminutif, jeu</i>
<i>préfixe</i> _{tsi-}	RADICAL _{dupl} → X _{tsi}	<i>diminutif, jeu</i>
<i>préfixe</i> _{tsi-}	NUMÉRAL → Adv _{numéral}	<i>'n par n'</i>
<i>préfixe</i> _{in-}	NUMÉRAL → Adv _{numéral}	<i>'n fois'</i>

<i>préfixe an-</i>	N → Adv _{loc}	<i>locatif</i>
--------------------	------------------------	----------------

Figure 6 : Fonctions morphologiques typées et fonctions sémantiques en malgache

Ce tableau résume l'essentiel de la morphologie dérivationnelle (et une partie de la morphologie flexionnelle⁴⁵) malgache. La première colonne décrit la fonction morphologique (le type d'affixe ou de processus dont il s'agit), la deuxième colonne identifie le type d'input et d'output de la fonction morphologique⁴⁶ et la troisième colonne donne une brève identification la fonction sémantique associée⁴⁷.

Lors de la création d'une base lexicale, il suffit donc que l'agent lexicographe (humain ou non) applique cette table et vérifie systématiquement lesquelles de ces relations sont effectivement réalisées, quelles sont leurs propriétés particulières et, éventuellement, quelles autres relations plus idiosyncratiques existent pour la racine en question⁴⁸.

⁴⁵ La morphologie flexionnelle est très limitée en malgache. Il n'y a pas de marque de personne, ni de nombre sur le verbe. Les modes se réduisent à deux, l'infinitif et l'impératif. Il n'y a pas de genre ni de nombre sur le nom ou l'adjectif et donc pas d'accord. Par contre les prépositions peuvent être tensées et il existe un système complexe de prépositions et d'adverbes locatifs.

⁴⁶ RADICAL et RADICAL_{thème} font référence aux formes de base. Le RADICAL_{thème} est une forme, contenant généralement une consonne thématique qui n'est pas totalement prévisible et qui apparaît devant suffixe (i.e., à l'impératif, au passif suffixé et au circonstanciel). RADICAL_{dupl} est la catégorie des radicaux redoublés. Pour le RADICAL_{varotra}, le RADICAL_{thème} est *varot-* et le RADICAL_{dupl} est *varobarotra*.

⁴⁷ Nous n'indiquons pas ici les types sémantiques de l'input et de l'output des fonctions sémantiques, puisque ce travail est en cours, mais le nom de la fonction sémantique constitue une indication grossière du type de son output. On peut rapprocher ces fonctions sémantiques des fonctions lexicales de la lexicologie explicative et combinatoire. Cependant, pour la lexicologie malgache, il nous semble plus intéressant de construire d'abord des fonctions morphologiques et de les associer à des fonctions sémantiques, plutôt que l'inverse, puisqu'une bonne partie de la morphologie dérivationnelle est très régulière (et peut donc être générée automatiquement et vérifiée de façon efficace). Les types morphologiques sont des catégories indicées. Comme il s'agit d'une représentation par contraintes, on peut poser des contraintes négatives sur ces indices. Par exemple, la spécification V^*ina/ana sur les préfixes *m-* et *mp-* indique qu'ils peuvent s'appliquer à n'importe quelle forme verbale sauf celles à suffixe (ou circonfixe contenant) *-ina* ou *-ana*. Notons aussi que l'on peut considérer les circonfixes malgaches comme une combinaison d'un préfixe et d'un suffixe. Dans ces conditions, on aurait les spécifications suivantes pour le suffixe *-ana* :

<i>suffixe -ana</i>	RADICAL _{thème} → V _{ana}	<i>passif</i>
<i>suffixe -ana</i>	V _{i/a} → V _{i/a-ana}	<i>circonstanciel</i>

⁴⁸ Il existe aussi quelques infixes fossiles à productivité très limitée, mais qui seront tout de même inclus dans nos bases lexicales.

Nous avons effectivement créé un environnement expérimental permettant de faciliter ce travail. Pour une forme donnée, les fonctions de la table sont automatiquement appliquées. L'agent lexicographe n'a plus qu'à compléter l'information (essentiellement sémantique) laissée implicite et à filtrer les données incorrectes. Les entrées correspondantes sont automatiquement créées, avec l'information partielle régulière qu'elles contiennent et associées à la racine par des hyperliens étiquetés (par les fonctions morphologiques et sémantiques) et bidirectionnels (les liens inverses étant évidemment étiquetés par les fonctions inverses). Les entrées résiduelles (i.e. ne correspondant pas à des formes effectives) peuvent être filtrées (automatiquement ou non)⁴⁹.

Si la base est créée exclusivement à partir d'institutions lexicales (i.e., sans tenir compte des connaissances linguistiques de l'agent et sans recours à des corpus externes), on peut tout de même obtenir des résultats intéressants. Ainsi, à partir de l'analyse des entrées de *varotra* dans les trois institutions lexicales A&M, RR et NR, on peut obtenir l'entrée complexe dont une représentation externe est donnée en annexe.

ANNOTATIONS

Les ANNOTATIONS sont très nombreuses⁵⁰ et peuvent être de plusieurs types dans les dictionnaires traditionnels malgaches.

D'abord, un type particulier : le renvoi éventuel à une RACINE. La plupart des entrées lexicales incomplètes et certaines entrées lexicales complètes contiennent une telle rubrique.

Par exemple :

« MARÀRY , Voy. RARY.»	(A&M, p. 434. Il s'agit du premier homonyme <i>rary</i> , 'mal, douleur', <i>marary</i> , 'malade, souffrant')
« Gija a. : Jer. Geja.»	(RR, p. 392. <i>Geja</i> 'serrer, étreindre, retenir')
« lahàsa V. asa.»	(NR, p. 180. Il y a quatre entrées homonymes <i>asa</i> . Il s'agit ici de <i>asa</i> ₂ , 'travail', dont <i>lahasa</i> ,

<i>infixe-in-</i>	RADICAL → V _{in}	<i>passif</i>
<i>infixe-om-</i>	RADICAL → V _{om}	?

Le premier permet de relier *vinidy* ('acheté') à *vidy* ('achat, prix') et le second *homehy* ('riant') à *hehy* ('rire (nom)'). Il y a également d'autres préfixes fossiles identifiables formellement mais non sémantiquement : *bo-*, *fo-*, *kan-*, *ko-*, *mo-*, *sa-*, *san-*, *tan-*, *to-*, *to-*, *tsin-*, *vo-*, *za-*. D'autres infixes servent à former des dérivés sur des catégories mineures (*-re-*, pluriel, *-za-*, lieu caché, *-n-*, locatif). Les fonctions morphologiques correspondantes existent, mais sans fonctions sémantiques associées. Elles peuvent être activées au besoin. Ainsi *bohesika* ('action d'entrer en masse') peut être traité comme un RADICAL simple ou comme un RADICAL dérivé de *sesika* ('bourre, bourrage').

⁴⁹ Les fonctions correspondantes retourneraient, par exemple, la liste vide.

⁵⁰ Ainsi, NR contient plus de 2 500 annotations pour un peu plus de 4 500 entrées.

qui n'apparaît pas dans l'entrée en question, est plus ou moins synonyme — plutôt spécialisé comme 'travail des champs' —).

Les entrées lexicales complètes qui contiennent un renvoi RACINE sont celles qui ne sont pas elles-mêmes des racines et dont la relation avec la racine n'est pas graphiquement transparente.

Par exemple :

«**KAVÏA** (*havia*), *adj.* Gaucher. **Kaviavia**, *adj.* Maladroit, lourd. **Mi—**, *vn*, Être maladroit. *Mikaviavia* izaitsizy izy, raha mamankona. Il est très maladroit pour raboter. **FI—**, *s.* La maladresse. **MPI—**, *s.* Un maladroit. **Fikaviaviàzana**, *s.* La maladresse. **Ikaviaviàazana**, *rel.*»

(A&M, p. 319. *Ka-* n'est pas un préfixe dérivationnel régulier.)

«**Maneza** (*heza*) *mt. p.* (*manezà*) : Manisy heza.»

(RR, p. 617. *Manisy heza* 'appliquer le heza', le heza est (selon A&M, p. 243) une «Eau extraite des cendres de certaines herbes et servant de mordant dans la teinture de la soie.» *heza* est la racine, *mt. p.*, la catégorie *matoantenin'ny mpanao*, 'litt. verbe d'acteur', *manezà*, la forme de l'impératif)

«**nahàndro** (*handro*) *n.* Mets, nourriture.»

(NR, p. 213. *Na-* n'est pas un préfixe dérivationnel régulier. Littéralement, il s'agit de la forme flexionnelle du passé de *mahandro*, 'cuisiner'.)

Dans une base lexicale, les liens sont établis systématiquement pour toutes les formes vers leurs racines, que les relations soient transparentes ou pas. On ne conservera donc pas ces renvois spécifiques.

Les autres renvois dans les dictionnaires traditionnels concernent des synonymes, des antonymes, des variantes, des abréviations, des surnoms, des formes supplétives (par exemple *amidy* comme forme passive de *varotra* dans A&M), le renvoi à des illustrations (dans RR), etc. Enfin, un certain nombre d'annotations constituent des marques d'usage. Dans nos bases lexicales, nous conservons ces annotations classées par type, en attendant de les distribuer et de les compléter de façon plus systématique.

4. CONCLUSION

Dans cet article, nous avons tenté de montrer comment on pouvait combiner de l'information lexicale provenant d'institutions lexicales distinctes dans une base lexicale dynamique. Nous avons examiné la structure des entrées lexicales dans trois institutions lexicales : Abinal & Malzac (1888), Rajemisa-Raolison (1985) et Rajaonarimanana (1995). Les dictionnaires traditionnels malgaches sont particulièrement intéressants à cet égard, puisqu'ils sont très riches et généralement bien structurés. De plus, comme la langue malgache possède une morphologie dérivationnelle complexe, mais très régulière, il est possible d'utiliser ces régularités pour filtrer et compléter les dictionnaires traditionnels. Le but à long terme de cette recherche est de favoriser l'émergence d'une véritable lexicographie expérimentale qui utilise au maximum les données des institutions lexicales traditionnelles.

RÉFÉRENCES

- ABINAL & MALZAC, SJ. (1888) : *Dictionnaire malgache-français*, Fianarantsoa, Ambozontany.
- BLACHE, P. & J.-Y. MORIN (1997) : «TAO et théories linguistiques : institutions grammaticales», communication aux V^{es} Journées scientifiques du réseau LTT, *La mémoire des mots*, Tunis, septembre 1997.
- BLANK, I. (1995) : «Sentence alignment : Methods and implementations», *T.A.L.*, 36, 1-2, pp. 81-99.
- CHOMSKY, N. (1995) : «Bare phrase structure», dans Webelbuth, G. (1995 éd.), *Government and Binding Theory and the Minimalist Program*, Oxford, Blackwell, pp. 383-439.
- DOMENICHINI-RAMIARAMANANA, B. (1983) : *Du ohabolana au hainteny, langue, littérature et politique à Madagascar*, Paris, Karthala.
- EVANS, R. & G. GAZDAR (1996) : «DATR : A language for lexical knowledge representation», *Computational Linguistics*, Vol. 22, 2, pp. 167-216.
- GALE, W. & K. CHURCH (1993) : «A program for aligning sentences in bilingual corpora», *Computational Linguistics*, 19, 1, pp. 121-142.
- MEL'CUK, I. et al. (1984, 1988, 1992) : *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques*, 3 volumes parus, Montréal, Presses de l'Université de Montréal.
- MEL'CUK, I., CLAS, A. et A. POLGUÈRE (1995) : *Introduction à la lexicologie explicative et combinatoire*, Montréal, AUPELF-UREF, Gembloux, Duculot.
- MILIÇEVIÇ, J. (1997) : *Étiquettes sémantiques dans un dictionnaire formalisé du type Dictionnaire Explicatif et Combinatoire*, Mémoire de maîtrise, Département de linguistique et de traduction, Université de Montréal.
- MORIN, Jean-Yves (1997) : «Lexiques-grammaires et institutions lexicales», communication aux V^{es} Journées scientifiques du réseau LTT, *La mémoire des mots*, Tunis, septembre 1997.
- PROFITA, P. (1969) : *Rakibolana malagasy-italianina/Dizionario italiano malgascio*, Antananarivo, P. Pietro Profita.
- PUSTEJOVSKY, J. (1995) : *The Generative Lexicon*, Cambridge, Mass., MIT Press.
- RABENILAINA, R.-B. (1987) : *Lexique-grammaire du malgache. Constructions transitives et intransitives*, Antananarivo, FOFIPA.
- RABENILAINA, R.-B. (1996) : *Le verbe malgache. Constructions transitives et intransitives*, Montréal, GRESLET et AUPELF-UREF.
- RAJAONARIMANANA, Narivelo (1995) : *Dictionnaire du malgache contemporain*, Paris, Karthala.
- RAKOTONAIVO, F. (1996) : *Rakibolana frantsay-malagasy*, Fianarantsoa, Ambozontany.
- SILBERZTEIN, M. (1993) : *Dictionnaires électroniques et analyse automatique de textes*, Paris, Masson.

ANNEXE : Une représentation externe du résultat de la fusion des trois entrées de *varotra*⁵¹

VEDETTE **varotra**

CAT n

TRADUCTIONS

Vente, commerce, marchandises NR
Vente, commerce, négoce, trafic, marchandises AM

DÉRIVÉS

DÉRIVÉ

TYPE MORPHO Duplicatif
 VEDETTE **varobarotra**
 CAT n
 TRADUCTIONS [*dupl. dim. de varotra*] AM
 AM RR NR

DÉRIVÉ

TYPE MORPHO [Circonfixe **fi-
-ana**]
 VEDETTE **fivarotana**
 CAT n
 TRADUCTIONS
L'action de vendre, de faire le commerce, de trafiquer, le prix, le lieu, la cause AM
 commerce NR
 EXPRESSIONS
 [MALGACHE **Fivarotamboky**]

⁵¹ Pour des raisons de concision, nous n'avons pas inclus dans cette représentation les définitions en malgache de RR, ni les champs ne contenant aucune valeur spécifiée positivement, ni non plus les fonctions sémantiques. Les abréviations AM NR et RR indiquent l'origine d'une information. Leur portée, implicite ici, est explicitement spécifiée dans les représentations internes. Cet exemple de représentation externe réduite est générée automatiquement à partir de la représentation interne. On pourrait aussi générer une représentation HTML, par exemple. La représentation interne est, une représentation prolog et contient beaucoup plus d'informations.

FRANÇAIS *librairie*
 MALGACHE **Fivarotampanafody**
 FRANÇAIS *pharmacie*
 MALGACHE **Fivarotan-tena**
 FRANÇAIS *prostitution*
 NR
 AM RR NR

DÉRIVÉ

TYPE MORPHO [Circonfixe **i--ana**]
 VEDETTE **ivarotana**
 CAT v_circ
 EXEMPLES
 MALGACHE **Ariary telopolo no nivarotako ny omby mifahiko.**
 FRANÇAIS *J'ai vendu trente piastres le boeuf que j'avais engraisé.* AM
 NOTES [IMPÉRATIF **ivaròty**] AM
 AM RR

DÉRIVÉ

TYPE MORPHO [Préfixe **fi-**]
 VEDETTE **fivarotra**
 CAT n
 TRADUCTIONS
Ce qu'on peut vendre, manière de vendre AM
 AM RR

DÉRIVÉ

TYPE MORPHO [Préfixe **maha-**]
 VEDETTE **mahavarotra**
 CAT v_abil
 TRADUCTIONS *Qui peut ou sait vendre* AM
 AM RR

DÉRIVÉ

TYPE MORPHO [Préfixe **mampi-**]
 VEDETTE **mampivarotra**
 CAT v_caus
 AM RR

DÉRIVÉ

TYPE MORPHO [Préfixe **mi-**]

VELETTE **mivarotra**
 CAT v_actif
 TRADUCTIONS
Vendre, faire le commerce,
trafiquer AM
Vendre, trafiquer NR
 EXPRESSIONS
 MALGACHE **Mivaro-**
tena RR
 NOTES [IMPÉRATIF
mivaròta] AM
 AM RR NR

DÉRIVÉ
 TYPE MORPHO [Préfixe
mifampi-]
 VELETTE **mifampivarotra**
 CAT v_récip
 AM RR

DÉRIVÉ
 TYPE MORPHO [Préfixe
mpi-]
 VELETTE **mpivarotra**
 CAT n
 TRADUCTIONS *Celui qui vend, le*
marchand, le commerçant AM
commerçant, marchand NR
 EXPRESSIONS
 MALGACHE **Mpivaro-**
kena
 FRANÇAIS *boucher*
 MALGACHE **Mpivaro-**
tena
 FRANÇAIS *prostituée*
 MALGACHE **Mpivaro-**
mandeha
 FRANÇAIS *marchand*
ambulant
 MALGACHE
Mpivarotra amoron-
dalana
 FRANÇAIS *marchand du bord des*
rues
 MALGACHE **Mpivarotra**
omby
 FRANÇAIS *marchand de*
bestiaux NR
 AM RR NR

DÉRIVÉ
 TYPE MORPHO [Préfixe **voa-**]
 VELETTE **voavarotra**
 CAT v_passif
 RR

EXPRESSIONS
 MALGACHE **Fifanakalozam-**
barotra
 FRANÇAIS *échanges commerciaux*
 NR
 MALGACHE **Fivarotam-boky**
 FRANÇAIS *librairie* NR
 MALGACHE **Fivarotam-panafody**
 FRANÇAIS *pharmacie* NR
 MALGACHE **Fivarotan-tena**
 FRANÇAIS *prostitution* NR
 MALGACHE **Lalam-barotra**
 FRANÇAIS *débouché* NR
 MALGACHE **Mamola-barotra**
 FRANÇAIS *Chercher à faire baisser, à*
réduire le prix AM
 MALGACHE **Mamola-barotra**
 RR
 MALGACHE **Manongom-barotra**
 [CF Songona] AM
 AM RR
 MALGACHE **Tombom-barotra**
 FRANÇAIS *Profit, bénéfice* AM
 FRANÇAIS *bénéfice* NR
 MALGACHE **Tranom-barotra**
 FRANÇAIS *maison, société de*
commerce NR
 MALGACHE **Varo-babo**
 FRANÇAIS *Très bon marché, à vil*
prix AM
 MALGACHE **Varo-balongana**
 [CF Valongana] AM
 MALGACHE **Varo-baventy**
 FRANÇAIS *Commerce en grand*
 [FIGURÉ *tout vol où on expose sa vie*]
 AM
 MALGACHE **Varo-bazaha**
 FRANÇAIS *Vente à prix fixe* AM
 AM RR
 MALGACHE **Varo-boba**
 [CF Varo-babo] AM
 FRANÇAIS *vente à vil prix, solde*
 NR
 AM RR NR
 MALGACHE **Varo-dratsy** RR
 MALGACHE **Varo-droa**
 FRANÇAIS *Action de vendre encore*
ce qu'on a vendu, de promettre à l'un ce qu'on
a promis à l'autre AM
 MALGACHE **Varodroa** RR
 MALGACHE **Varo-mahery**
 FRANÇAIS *Vente forcée* AM
 MALGACHE **Varo-maika**

FRANÇAIS *Vente faite à vil prix pour cause d'urgence*
 AM
 FRANÇAIS *vente d'urgence* NR
 AM RR NR
 MALGACHE **Varo-maizina**
 FRANÇAIS *marché noir* NR
 RR NR
 MALGACHE **Varo-mamindro**
 FRANÇAIS *Achat dont on fait traîner la discussion dans l'espérance de trouver mieux* AM
 MALGACHE **Varo-mandeha**
 FRANÇAIS *Colportage* AM
 NR
 AM RR NR
 MALGACHE **Varo-maty**
 FRANÇAIS *Vente définitive* AM NR
 AM RR NR
 MALGACHE **Varo-miandry**
 FRANÇAIS *Objets qu'on garde longtemps, qui ne se vendent que par circonstance, dont on demande un bon prix*
 AM
 AM RR
 MALGACHE **Varo-miera**
 FRANÇAIS *Vente d'un article qui peut être rendu si l'acquéreur n'en est pas content; action de prendre à crédit avec obligation de rendre l'objet à défaut du prix* AM
 AM RR
 MALGACHE **Varo-mifody** RR
 MALGACHE **Vodiondrim-barotra**
 FRANÇAIS *arrhes* NR
 MALGACHE **Ady varotra**
 FRANÇAIS *marchandage, négociation* NR
 RR NR
 MALGACHE **Miady varotra**
 FRANÇAIS *marchander, négociier* NR
 MALGACHE **Varo-mahay mody**
 FRANÇAIS *Vente d'un article qui peut être rendu si l'acquéreur n'en est pas content; action de prendre à crédit avec obligation de rendre l'objet à défaut du prix* AM
 RR
 MALGACHE **Varotra mahay mody**
 RR
 MALGACHE **Varo-tsy azon-dahy**
 FRANÇAIS *Marché sur lequel on ne s'entend pas, malentendu dans une vente ou un achat* AM
 MALGACHE **Varo-tsy mifody**
 [LITTÉRALEMENT *Vente sur laquelle on ne revient pas*]
 [FIGURÉ *la mort*] AM

AM RR
 MALGACHE **Varotsimifody**
 FRANÇAIS *l'au-delà, l'autre monde*
 [LITTÉRALEMENT *vente sur laquelle on ne revient pas*] NR
 MALGACHE
Varotr'omby anaty ambiaty
 [LITTÉRALEMENT
Vente de boeufs cachés dans les ambiaty]
 [FIGURÉ *vente malhonnête*] AM
 MALGACHE
Varotra omby anaty ambiaty
 RR
 MALGACHE **Varotra alika** RR
 MALGACHE **Varotra**
ambongadiny
 FRANÇAIS *Vente en gros, par pièce*
 AM
 FRANÇAIS *vente en gros* NR
 AM RR NR
 MALGACHE **Varotra an'elakelatrano**
 RR
 MALGACHE **Varotra an-trano**
tokana
 FRANÇAIS *Monopole* AM
 AM RR
 MALGACHE **Varotra antsinjarany** AM
 MALGACHE **Varotra antsinjarany**
 RR NR
 FRANÇAIS *Vente au détail*
 AM RR NR
 MALGACHE **Varotra**
fampisehoana
 FRANÇAIS *exposition-vente* NR
 MALGACHE **Varotra iombonana**
 FRANÇAIS *communauté économique*
 NR
 MALGACHE **Varotra ivelany**
 FRANÇAIS *commerce extérieur*
 NR
 MALGACHE **Varotra misy fehivavany** RR
 MALGACHE **Varotra olona** RR
 MALGACHE
Varotra tsy
miverimbodiondry RR
 MALGACHE **Varotry ny maika hiala** RR
 MALGACHE **Varotry ny maika hody** RR
 MALGACHE **Varotry ny mananotena**

[LITTÉRALEMENT *Commerce des veuves*]

[FIGURÉ *vente faite avec hésitation, dans la crainte de se tromper*] AM

AM RR

MALGACHE **Varotra ankobohobo** RR

MALGACHE

Varo-jabo an-tsena, ka samy mandoka ny azy ho masaka.

FRANÇAIS (prov.) *Chacun vante sa marchandise, comme les marchands de Jabo au marché.* AM

EXEMPLES

MALGACHE

Aza hitsahinao ity varotr'olona. RR

MALGACHE

Lasa nanao varotra any antsiraka izy.

RR

NOTES

[PASSIF Amidy] [CF Vidy]

AM

ÉTUDE POUR UNE EXTRACTION AUTOMATIQUE DE NÉOLOGISMES

Yvette Yannick MATHIEU

CNRS, Laboratoire de Linguistique Informatique, Université Paris 13, France

Le but de cette étude est le repérage automatique de néologismes de forme dans des textes écrits. Nous avons comparé, par ordinateur, le vocabulaire d'une année d'un journal quotidien (*Le Monde* 1993) avec la nomenclature de l'année 1992 du dictionnaire électronique *DELAF*¹ qui contient les formes fléchies des mots simples. Ce dictionnaire a été constitué au Laboratoire d'Automatique Documentaire et Linguistique (LADL) par Blandine Courtois (1990). Les résultats de cette comparaison ont été analysés afin de définir des algorithmes de traitements informatiques pour discriminer les erreurs et reconnaître les néologismes potentiels.

Cet article complète celui de Mathieu, Gross et Fouqueré (1998) dans lequel nous avons étudié en détail les mots commençant par les lettres **a**, **b** et **c**. L'étude que nous présentons ici est maintenant exhaustive puisque nos résultats concernent le corpus entier; précisons également qu'un dictionnaire de référence différent a été utilisé².

1. RÉSULTATS

Le corpus du journal *Le Monde* de l'année 1993 contient 23 millions de chaînes de caractères³ dont 201 775 différentes. Chaque chaîne a été recherchée dans le dictionnaire, celles qui en sont absentes sont des noms propres, des mots avec tiret ou des mots simples.

Il est naturel de ne trouver aucun mot avec tiret dans le *DELAF*, car ce dictionnaire ne contient que des mots simples.

Nous avons dénombré 123 044 noms propres. Nous appelons «nom propre» tout mot qui commence par une lettre majuscule, qui n'est pas un sigle et qui n'est pas le début d'une phrase, c'est-à-dire qui n'a pas été reconnu dans sa graphie commençant par une

¹ Dictionnaire Electronique du LADL des Formes fléchies.

² La problématique du choix du dictionnaire de référence ne sera pas abordée ici.

³ Nous appelons chaîne de caractères toute séquence de lettres contiguës contenant éventuellement un ou plusieurs traits d'union mais ni blanc ni signe de ponctuation. Les chaînes en majuscules et minuscules sont distinguées.

minuscule. Ces mots représentent la grande majorité des mots inconnus dans les corpus de journaux. Max Silberztein (1995) a fait une étude comparative des corpus du *Monde* 1992 et 1993 et a dénombré 89 % de mots en majuscules parmi les mots absents du *DELAFA*.

Il y a 15 230 mots simples du *Monde* 1993 qui ne sont pas reconnus par le *DELAFA*. Parmi ces mots, il y a des néologismes possibles et des erreurs de diverses natures.

1.1 Les néologismes possibles

Nous avons fait une étude précise de tout le corpus des mots simples qui peuvent être des néologismes. Nous pouvons distinguer :

- des noms et adjectifs dérivés de nom de ville, de région ou pays : *ajacciens, catalanisation, corrézien, corsitude, marocanisation, paimpolaise*;

- des noms et adjectifs dérivés de nom de personnalité artistique ou politique. Ces mots sont pour la plupart liés à l'actualité et leur emploi est probablement éphémère : *balladurien(ienne), balladurisation, balladurisé, balladurisme, chiraquien(ienne), chiraquie, chiraquisme, etc.* Mais il y a également des dérivés de personnages célèbres plus anciens : des compositeurs (*berlioziste, brahmsien*), des peintres (*caravagesque, chagallien, raphaëlesque*), des écrivains (*brechtien, brechtisme, conradienne, courtelinesque, dickensien, éluardien, perecquien*), des cinéastes (*fellinien, rohmérien*), des musiciens (*coltranien*). Ces mots dérivés, qui ne vont probablement pas disparaître de la langue, pourraient soit être intégrés au dictionnaire de référence, soit constituer un dictionnaire à part entière.

- des noms de vins : *bourgueil, margaux, yquem*;

- des abréviations ou diminutifs, utilisés dans le langage courant : *ados* (adolescents), *allocs* (allocations), *clodo* (clochard), *cata* (catastrophe), *séropo* (séropositif). Ces mots ont été étudiés par Groud et Serna (1996);

- des mots préfixés et des formes en *o-* qui sont des réductions d'adjectif ou de nom coordonné (Dugas, Mathieu et Courtois 1995). Certains préfixes sont très fréquents : *anti* (*antibourgeois, anticapitaliste, antihéros*), *co* (*coanimateur, codécouvreur, coprésence*), *dé* (*décohabitations, déprivatisées, déréférencer*);

Le tableau suivant indique le nombre d'occurrences des préfixes les plus courants.

Préfixes	Nombre
<i>anti</i>	275
<i>auto</i>	112
<i>co</i>	158
<i>dé</i>	114
<i>euro</i>	42
<i>hyper</i>	35
<i>inter</i>	83
<i>micro</i>	38
<i>multi</i>	69
<i>néo</i>	40
<i>pré</i>	83
<i>pro</i>	34
<i>re</i>	89
<i>ré</i>	57

Tableau 1

- des mots étrangers, essentiellement anglais, italiens, allemands et latins : *addict*, *network*, *nocciola*, *overseas*. Certains sont utilisés couramment dans la langue comme *bodybuilding*, *brunch*, *bulldog*, *fitness* (Girardin 1996) et produisent parfois des dérivés francisés : *bodybuildé*, *folkeux*;

- des mots nouveaux (c'est-à-dire qui ne sont pas dans le *DELAF*) qui ne font pas partie des catégories déjà mentionnées comme : *accessoirisé*, *accréditation*, *anabolisant*, *animalerie*, *audimat*, *automaticiens*, *avionique*, *crédibiliser*, *contractualiser*, *photocopillage*, *scénarisation*. Certains mots sont construits à partir d'initiales. Ainsi *bédéphile* vient de B.D. (Bandes Dessinées), *pédégères* vient de P.D.G. (Président Directeur Général). Des adjectifs sont créés à partir de ceux existants en ajoutant un suffixe pour amplifier ou modifier le trait : *brillantissime*, *élégantissime*, *kitchissime*, *lentissime*, *nullissime*, *jargonnesque*, *spiralesque*, *vampiresque*. Une intensification est également donnée par une orthographe différente : *hénaurme* pour *énorme*. Des adjectifs sont créés en ajoutant le suffixe *phobe* : *homophobe*, *motophobe*, *publiphobe*. Il existe une tendance à la féminisation de noms ou d'adjectifs : *baroudeuses*, *forçate*, *écrivaine*, *loubarde*.

Cette typologie est résumée dans le tableau 2.

Types	Nombre	% des néologismes potentiels
Néologismes potentiels	7170	
<i>N</i> et <i>Adj</i> dérivés de noms géographiques	855	11,9
<i>N</i> et <i>Adj</i> dérivés de noms de personne	370	5
Abréviations et diminutifs	170	2,37
<i>N</i> , <i>V</i> et <i>Adj</i> préfixés	2030	28,3
Autres mots simples	2280	31,8

Tableau 2

Parmi les mots simples inconnus du *DELAF*, il y a de nombreuses erreurs. Nous en avons fait également une typologie, afin de définir des algorithmes pour les éliminer automatiquement par ordinateur.

1.2 Les erreurs

Les erreurs sont principalement typographiques : mots collés comme : *leministre*, *lacrétation*, *étédécidées*, *catastrophiquedes*, *françaiset* (1 340 en tout), mots coupés : *onneur*, *sec*, *uions*, et des mots mal orthographiés, la frontière entre ces deux types d'erreur n'étant pas clairement délimitée. (Voir Mathieu et *alii.* (1997) pour une étude détaillée.)

2. STRATÉGIE POUR RECONNAÎTRE LES «CANDIDATS-NÉOLOGISMES»

L'étude précédente permet de définir les différentes étapes d'un traitement informatique. Pour établir si un mot inconnu du dictionnaire de référence peut être un néologisme, nous proposons :

1) d'isoler les mots simples;

2) d'éliminer les erreurs : les lettres isolées, les mots qui ont 3 consonnes identiques qui se suivent, les mots collés, les mots qui ont des erreurs d'accent (cette dernière étape en utilisant un correcteur orthographique);

3) d'éliminer les mots qui présentent une malformation régulière comme *arborigène* et les mots tronqués, par consultation de dictionnaires préalablement établis.

Une analyse du corpus restant doit permettre de reconnaître les mots qui peuvent être des néologismes. Un certain nombre d'opérations peuvent être exécutées, comme :

- extraire les mots préfixés, avec une méthode que nous avons décrite dans Mathieu et *alii.* (1997);

- extraire les mots contenant certains suffixes comme *esque* ou *issime*;

- repérer les mots issus de noms géographiques ou de noms propres. Ce repérage pourrait se faire grâce à une étude sur les processus de dérivation de ces noms. Denis Maurel (1994) a commencé ce travail pour les noms de géographie;

- reconnaître certaines catégories de mots, par consultation de dictionnaires précédemment établis : les mots étrangers, les noms de vins, etc.

La démarche que nous venons de décrire est rendue possible, d'une part, par la disponibilité de grands corpus de textes et, d'autre part, par l'existence ou le développement d'importantes ressources linguistiques électroniques comme les dictionnaires de formes fléchies, les dictionnaires géographiques, etc.

CONCLUSION

Cette étude s'inscrit dans un courant de recherche qui privilégie l'étude de phénomènes linguistiques à travers de grands corpus. Des travaux se sont plus particulièrement intéressés à la détection de néologismes dans la presse (voir, par exemple, Cabré et de Yzaguirre (1995) pour le catalan et le castillan). L'objectif principal de notre travail est d'alléger considérablement la tâche du linguiste, en lui proposant des procédures semi-automatiques pour constituer un corpus de néologismes éventuels. Mais il reste néanmoins un travail important à faire : établir si un néologisme potentiel est ou non un vrai néologisme.

RÉFÉRENCES

- CABRÉ, Maria Teresa et Luis de YZAGUIRRE (1995) : «Stratégie pour la détection semi-automatique des néologismes de presse», *Traduction Terminologie Rédaction*, vol. VIII, n° 2, Canada, Université Concordia, pp. 89-100.
- COURTOIS, Blandine (1990) : «Un système de dictionnaires électroniques pour les mots simples du français», *Langue française*, 87, Paris, Larousse.
- DUGAS, André, Yvette Yannick MATHIEU et Blandine COURTOIS (1995) : «Formes simples en o-», *Linguisticae Investigationes*, vol. XIX, n° 2, Amsterdam / Philadelphia, John Benjamins, pp. 429-441.
- GIRARDIN, Chantalle (1996) : «Trader : aux frontières du néologisme», *Terminologies nouvelles*, Rint (Réseau international de néologie et terminologie), n° 14, Ministère de la Culture et des Affaires sociales de Belgique, Bruxelles, pp. 31-34.
- GROUD, Claudette et Nicole SERNA (1996) : *De abdom à zoo, Regards sur la troncation en français contemporain*, CNRS-INaLF, Paris, Didier Érudition, 161 pages.
- MATHIEU, Yvette Yannick, Gaston GROSS et Christophe FOUQUERÉ (1998) : «Vers une extraction automatique des néologismes», *Cahiers de lexicologie*, Paris, Didier Érudition.
- MAUREL, Denis (1994) : «Le traitement informatique de la dérivation des noms de ville», *T.A.L.*, vol. 35, n° 2, Paris, ATALA, pp. 111-127.
- SILBERZTEIN, Max (1995) : «Dictionnaires électroniques et comptage de mots», *JADT*, Rome, pp. 93-101.

REPÈRES MÉMORIELS DU LEXIQUE DANS LES LANGUES BANTU. ILLUSTRATION À PARTIR DES STRUCTURES MORPHO-SÉMANTIQUES DES SUBSTANTIFS KIRUÚNDI

Jean B. NTAKIRUTIMANA

GRESLET, Université de Montréal, Canada

INTRODUCTION

Le groupe bantu représente un très large ensemble de langues parlées en Afrique centrale, orientale et australe. Globalement, la zone bantu s'étend au sud d'une ligne reliant le sud du Nigeria au sud de la Somalie (cf. carte en annexe).

Les langues bantu sont des langues remarquablement homogènes quasiment à tous les niveaux linguistiques, tellement homogènes que toutes les études historico-comparatives leur attribuent aisément et unanimement une maternité linguistique commune, la proto-langue ayant été reconstituée à la fin du XIX^e siècle par Meinhof (1899) qui l'a nommée Ur-Bantu¹.

Les langues appartenant à la famille bantu possèdent une structure morpho-lexicale particulière par sa grande régularité. Cette régularité facilite une grande prédictibilité de la valeur sémantique de chaque mot, ainsi qu'une grande souplesse dans la créativité lexicoterminologique; ce qui, comme nous le montrerons, amplifie considérablement la mémoire linguistique de chaque mot.

Le kiruúndi appartient au domaine linguistique bantu². Sa structure linguistique correspond au modèle typique des structures linguistiques bantu, c'est pourquoi nous partirons du lexique kiruúndi, particulièrement du substantif, pour illustrer nos propos sur les repères mémoriels du lexique bantu.

Nous présenterons, dans cette communication, les principaux éléments morphologiques du lexique kiruúndi où s'ancre préférentiellement la mémoire linguistique, ainsi que l'interactivité de ces repères d'ancrage au niveau intra et extra-lexical. Nous en

¹ L'Ur-Bantu est également appelé Proto-Bantu ou bantu commun.

² Le kiruúndi fait partie du grand groupe I.A.5., selon la classification de Greenberg (1966); il porte le numéro d'identification D62 dans la classification de Guthrie (1971), et le numéro J62 selon Meeussen (1967).

montrerons, par la suite, les corrélations avec les structures morphologiques types des autres langues bantou.

Nous proposerons également une façon d'organiser et de gérer les données relatives à ces structures morphologiques et, enfin, nous présenterons les applications possibles de cet aspect de la mémoire des mots dans divers domaines linguistiques et paralinguistiques.

1. SUBSTANTIF KIRUÚNDI

1.1 Structure morphologique

Le nom³ kiruúndi est constitué de 6 types de morphèmes, à savoir :

1. l'augment
2. le préfixe
3. l'infixe
4. le radical
5. le suffixe
6. la finale

N.B. : L'ensemble formé par le radical, ses suffixes éventuels et la finale, est appelé thème.

Le schéma ci-dessous montre comment les morphèmes — encadrés — se combinent entre eux pour constituer un nom, à travers une série d'étapes strictement ordonnées, partant du point de départ 0 → vers le point d'arrivée → 0.

La flèche qui contourne un morphème, de gauche à droite, signifie que le morphème en question peut être présent ou non dans la structure substantivale. C'est le cas de tous les morphèmes nominaux, à l'exception du préfixe et du radical qui ne peuvent jamais manquer dans un nom.

La flèche qui contourne un morphème, de droite à gauche, signifie que l'on peut retrouver plus d'un morphème de cette catégorie au sein d'un même substantif, situation très fréquente pour les infixes et les suffixes.

La flèche qui contourne, de droite à gauche, une série de morphèmes, veut dire que la série peut être redupliquée à l'intérieur d'un même nom, comme on le verra en 1.2.2.

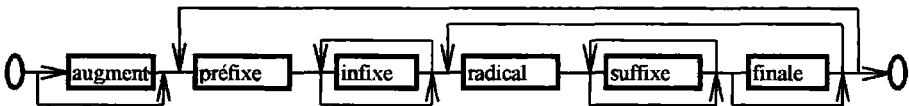


Figure 1 : Schéma de la structure morphologique du substantif kiruúndi

³ Ici, nous ne faisons aucune distinction entre les notions de «nom» et de «substantif».

Le mécanisme de construction substantivale schématisé ci-dessus est très productif en kiruúndi. Théoriquement, il est même capable de générer une infinité de noms différents, qu'aucun ouvrage ne pourrait inventorier au complet.

Le tableau suivant présente, à titre d'exemples, quelques-uns des mots construits sur le modèle du schéma précédent, tandis que les découpages morphologiques illustrent la variété et la souplesse de leurs combinaisons.

SUBSTANTIFS	STRUCTURES MORPHOLOGIQUES					
	Augment	Préfixe	Infixe	Radical	Suffixe	Finale
1. <i>Abatáhahára</i> ≡ 'qui sont toujours présents'	a-	-ba-	-ta- -ha-	-har-		-a
2. <i>Agakomére</i> ≡ 'petite blessure'	a-	-ka-		-kom-	-ir-	-e
3. <i>Agatéerantimba</i> ≡ 'source d'inquiétudes'	a- i- ⁴	-ka- -n-		-ta- -tiimb-	-ir-	-a -a
4. <i>Imigoondooro</i> ≡ 'pousses [végétales]'	i-	-mi-		-goond-	-uur-	-a
5. <i>Inkórabára</i> ≡ 'qui est très dangereux'	i- i-	-n- -i-		-kor- -bar-		-a -a
6. <i>Intóboro</i> ≡ 'trou'	i-	-n-		-tob-	-ur-	-o
7. <i>Ishúurwé</i> ≡ 'fleur'	i-	-i-		-shuur-	-u-	-e
8. <i>Ubugúfi</i> ≡ 'état de ce qui est court'	u-	-bu-		-gufi		
9. <i>Udukomére</i> ≡ 'petites blessures'	u-	-tu-		-kom-	-ir-	-e
10. <i>Ugutóoratoora</i> ≡ 'ramassage répété'	u-	-ku-		-ta- -ta-	-uur- -uur-	-a -a
11. <i>Umukóbekóbe</i> ≡ 'figus'	u-	-mu-		-kobe- -kobe		
12. <i>Umutázitéeraruungu</i> ≡ 'qui ne les [vaches] ennuie pas'	u- i-	-mu- -i-	-ta- -zi-	-ta- -ruung-	-ir-	-a -u
13. <i>Umutúuhúrana</i> ≡ 'qui n'induit pas en erreur'	u-	-mu-	-ta-	-uh-	-ur-an-	-a
14. <i>Umuuntu</i> ≡ 'personne [humaine]'	u-	-mu-		-ntu		
15. <i>Umwímenyerezo</i> ≡ 'exercice'	u-	-mu-	-ii-	-meny-	-ir- -ir- -i-	-o

⁴ La lettre-morphème barrée subit un amuïssement dans la structure de surface.

16. <i>Urukurúkurú</i> ≡ 'rumeur'	u-	-ru-		-kur- -kur-		-u -u
17. <i>Urushíingé</i> ≡ 'aiguille'	u-	-ru-		-shiing-		-e

Tableau 1 : Structure morphologique de quelques substantifs kiruúndi

On ne peut dénombrer tous les radicaux du kiruúndi, et il s'en crée, de temps à autre, de nouveaux pour combler les besoins de communication; ce qui fait que le nombre de radicaux est théoriquement illimité. Contrairement au radical, les autres morphèmes lexicaux sont en nombre limité. Ils constituent des listes fermées, faciles à inventorier.

Ainsi donc, la création lexicale se fait autour du même noyau radical, auquel sont adjoints les différents morphèmes illustrés par la figure ci-dessus. Tous ces morphèmes interagissent entre eux et ont chacun un rôle particulier à jouer dans la structure du mot. Chaque morphème possède une valeur linguistique propre, non autonome et incomplète, en interaction avec les valeurs linguistiques des autres morphèmes; tandis que la valeur linguistique de la totalité du mot créé dépend des valeurs combinées des morphèmes qui le constituent.

Ce mode de création est extrêmement productif et régulier, de façon que tout locuteur soit apte à créer de nouvelles unités lexicales. Il est aussi normalement capable de deviner le sens de tout mot, même nouveau, si la structure linguistique du mot est correcte, et si la mémoire du locuteur en reconnaît le sens du radical de base.

1.2 Repères mémoriels

Nous considérons que la valeur linguistique du mot correspond à sa charge mémorielle, laquelle permet que le mot soit inscrit dans la mémoire individuelle du locuteur ainsi que dans la mémoire collective de la société où ce mot s'emploie.

Bien entendu, à chaque niveau linguistique correspond une mémoire linguistique propre. Chaque langue est dotée en effet de ses propres mémoires phonologique, syllabique, morphématique, lexicale, propositionnelle, sémantique, etc. qui, respectivement, régissent la structuration, la génération, le stockage et la combinaison des phonèmes, des syllabes, des morphèmes, des mots, des propositions, de la charge sémantique des éléments linguistiques, etc.

Par ailleurs, les éléments de chaque niveau mémoriel restent en constante interaction entre eux et servent de base à l'élaboration du niveau mémoriel directement supérieur; les autres niveaux linguistiques servant de repères mémoriels supplémentaires dans le discours.

Or, comme les mots sont constitués de morphèmes, il devient utile de comprendre que la mémoire linguistique d'un mot se manifeste au niveau de chaque morphème lexical. Celui-ci fonctionne alors comme principal repère mémoriel pour guider la mémoire dans les processus mentaux de synthèse (construction) et d'analyse (compréhension) des unités lexicales.

Nous survolerons, ci-dessous, et à titre simplement illustratif, certains lieux d'interaction entre les principaux repères mémoriels lexicaux — à savoir les morphèmes — et les niveaux phonologique, lexical et sémantique, dans la structuration des mots.

1.2.1 Niveau phonologique

STRUCTURE PHONOLOGIQUE DE L'AUGMENT

L'augment, appelé encore prépréfixe par Lemaréchal (1989 : 41), est la voyelle initiale du substantif kiruúndi, toujours suivie du préfixe de classe. Elle ne peut être que l'une des 3 voyelles du triangle vocalique du kiruúndi⁵ — en l'occurrence [i], [a] ou [u] — et doit toujours être identique à la voyelle du préfixe.

(1)

[i]	<i>imigoondooro</i> (<i>i-mi-goond-uur-o</i>)
[a]	<i>agatéerantúmba</i> (<i>a-ka-ta-ir-a-i-n-tiimb-a</i>)
[u]	<i>umuuntu</i> (<i>u-mu-ntu</i>)

Ainsi, la structure de l'augment prend pour repère principal la structure vocalique du préfixe de classe.

L'augment est la voyelle caractéristique de la catégorie substantivale où on la retrouve exclusivement⁶.

C'est également une voyelle qui subit de fréquents amuïssements, notamment quand le substantif est à la forme appellative, lorsqu'il est précédé d'un référent, d'une négation, etc., et surtout lors des créations lexicales par combinaison de plusieurs mots (cf. tableau 1, colonne 2).

LOI DE DAHL

Il a été observé que, dans le discours, les consonnes sourdes des affixes précédant le radical deviennent sonores lorsque la consonne initiale du radical est sourde. Ce phénomène de dissimilation consonantique, particulièrement fréquent dans les langues bantu de la région orientale de l'Afrique, a été baptisé «loi de Dahl» par les bantuistes. En kiruúndi, il concerne les consonnes *k* et *t* — les seules consonnes non voisées des préfixes de classes — qui se transforment respectivement en *g* et *d*.

(2)

<i>Udukomére</i>	< <i>u-tu-kom-ir-e</i>
<i>Agatéerantúmba</i>	< <i>a-ka-ta-ir-a-i-n-tiimb-a</i>

⁵ Le kiruúndi a 5 voyelles, ainsi disposées dans le triangle vocalique :

[i]		[u]
	[e]	[o]
	[a]	

⁶ C'est ce qui a fait croire, à tort, à certains linguistes, que l'augment correspondait à l'article en kiruúndi.

Les exemples ci-dessus montrent ainsi comment le radical sert de repère pour la structuration phonologique des morphèmes qui le précèdent, justifiant ainsi les rapports d'interaction énoncés ci-haut.

HARMONIE VOCALIQUE

Bien souvent, des suffixes changent de voyelles selon le type de radical qui les précède. C'est le cas du suffixe *-ir* qui se métamorphose en *-er* sous l'influence du *o* du radical, dans les noms suivants :

- (3)
- | | | |
|-------------------|-------------------------|--------------------------|
| <i>Agatóorero</i> | < <i>a-ka-toor-ir-o</i> | < <i>a-ka-ta-ur-ir-o</i> |
| <i>Udukomére</i> | < <i>u-tu-kom-ir-e</i> | |

Ces changements sont dus à une assimilation vocalique du suffixe, conditionnée par la voyelle pénultième du radical, dans une optique de rapprochement des timbres des voyelles concernées. Le radical devient ainsi un repère pour la structure vocalique du suffixe, comme il l'était pour la structure consonantique du préfixe selon la loi de Dahl.

1.2.2 Niveau lexical

SUBSTANTIVATION PAR PRÉFIXATION ET PRÉPRÉFIXATION

Toute unité lexicale, de quelque nature qu'elle soit, est susceptible d'être substantivée par adjonction d'un préfixe de classe et d'un augment :

- (4)
- Ugutóoratoora* : Le verbe *gutóoratoora* (*ku-ta-ur-a-ta-ur-a*) se nominalise en troquant le préfixe de l'infinitif *gu-* contre le préfixe *-ku-* de classe 15, naturellement précédé du prépréfixe correspondant *u-*.
- Ubugúfi* : L'adjectif *-gúfi* se transforme en substantif par préfixation en *-bu-* (classe 14) et par prépréfixation en *u-*.

Ce mode de création lexicale permet de comprendre la portée de la charge mémorielle particulière du préfixe de classe et du prépréfixe correspondant, dans la structure lexicale.

REDOUBLEMENT THÉMATIQUE

Il existe un autre mode de création lexicale consistant à redoubler le thème, en vue d'ajouter un sens diminutif ou répétitif à la charge sémantique du radical :

- (5)
- urukurúkuru* : répétition du thème *-kurú* pour désigner une *rumeur* (*urukurúkuru*) relativement moins importante qu'une vraie *nouvelle* (*inkurú*).
- ugutóoratoora* : redoublement du thème *-toora* (*ramasser*), pour signifier l'action de ramasser à plusieurs reprises.

Ce mécanisme d'enrichissement lexical recourt donc aux mêmes repères mémoriels que la simple création lexicale, mais exploite doublement les repères entrant dans la constitution du thème nominal, en l'occurrence le radical, ses suffixes et la finale.

AGGLUTINATION MULTILEXICALE

En plus d'exploiter doublement certains morphèmes lexicaux pour créer de nouvelles unités lexicales, le kiruúndi, comme d'autres langues bantu, procède à la combinaison de plusieurs mots de radicaux et de sens différents pour créer de nouvelles unités lexicales à charge mémorielle plus dense :

(6)

Agatéerantiimba : agglutination de deux unités lexicales issues de deux thèmes, *-teera* et *-tiimba*

Umutázitéeraruungu : agglutination de deux unités lexicales issues de deux thèmes, *-teera* et *-ruungu*

Les mots formés par ce mode d'agglutination possèdent une charge mémorielle plus importante parce qu'issue de morphèmes de plus d'une unité lexicale; ce qui équivaut à des repères mémoriels plus nombreux et plus variés.

ACCORDS DE CLASSES

Au-delà des repères morphologiques au sein d'une même unité lexicale, le substantif kiruúndi peut exercer une certaine influence sur les mots en contact avec lui dans un énoncé. Par son influence sur les autres unités lexicales, le substantif transfère une partie de sa charge mémorielle en prenant toujours pour repère le préfixe de classe. Effectivement, dans n'importe quelle proposition kiruúndi, c'est la classe du préfixe nominal qui commande tous les accords et concordances des mots à morphologie variable, tels que le verbe, l'adjectif, le pronom, etc. Voici des exemples d'accord dans la phrase suivante :

(7)

abaantu bagúfi bagira udutáambwé tugúfi

/ a-**ba**-ntu / **ba**-gufi / **ba**-gir-a / u-**tu**-tamb-u-e / **tu**-gufi /

/ humains / courts / font / petits pas / courts /

'les gens de petite taille font de petits pas'

Cet énoncé illustre comment, d'une part, le préfixe **-ba-** (classe 2) de *abaantu* commande l'accord de l'adjectif *bagúfi* et celui du verbe *bagira*; tandis que d'autre part, le préfixe **-tu-** (classe 13) de *udutáambwé* commande l'accord de l'adjectif *tugúfi*, issu, tout comme *bagúfi*, du même adjectif *-gufi*, soumis à l'effet de deux classes différentes.

1.2.3 niveau sémantique

MÉMOIRE SÉMANTIQUE DE L'AUGMENT

À regarder de près la structure sémantique du substantif kiruúndi, il semble que l'augment ne représente pas d'intérêt particulier au niveau de la mémoire sémantique du mot.

Le peu d'intérêt sémantique que représente l'augment fait que son fréquent amuïssement ne nuise en rien à la valeur sémantique des unités lexicales concernées.

MÉMOIRE SÉMANTIQUE DES PRÉFIXES

Dans les langues bantu, le nombre de préfixes nominaux (classes) varie entre 5 et 20; le kiruúndi en compte 16, comme l'illustrent les exemples ci-dessous.

(8)

1 -mu-	ex. (u-mu-ntu) umuntu = un humain
2 -ba-	ex. (a-ba-ntu) abaantu = des humains
3 -mu-	ex. (u-mu-vumu) umuvumú = ficus
4 -mi-	ex. (i-mi-vumu) imivumú = ficus
5 i- (-ri-)	ex. (i-i-vi) iví = genou
6 -ma-	ex. (a-ma-vi) amaví = genoux
7 -ki-	ex. (i-ki-ntu) ikiintu = chose
8 -bi-	ex. (i-bi-ntu) ibiintu = choses
9 -n-	ex. (i-n-ka) inká = vache
10 -n-	ex. (i-n-ka) inká = vaches
11 -ru-	ex. (u-ru-ntu) uruuntu = grosse chose
12 -ka-	ex. (a-ka-ntu) akaantu = petite chose
13 -tu-	ex. (u-tu-ntu) utuuntu = petites choses
14 -bu-	ex. (u-bu-ntu) ubuuntu = personnalité
15 -ku-	ex. (u-ku-ntu) ukuuntu = manière
16 -ha-	ex. (a-ha-ntu) ahaantu = lieu

Ce dénombrement des préfixes de classes ne tient pas compte de l'opposition singulier/pluriel, car dans les habitudes linguistiques bantuistes, on considère qu'un mot au singulier appartient à une classe différente de celle du même mot au pluriel.

Par ailleurs, on remarque que des mots appartenant aux mêmes catégories sémantiques⁷ ont tendance à se regrouper au sein des mêmes classes.

Très souvent, en effet, les classes :

1/2	sont réservées aux humains;
3/4	regroupent beaucoup de noms de végétaux;
5/6	sont utilisées pour les parties du corps;
6	à elle seule désigne les liquides et les masses;
7/8	- se retrouvent dans beaucoup de noms d'animaux, - et donnent parfois un sens augmentatif;
9/10	regroupent également beaucoup de noms d'animaux;
11	est augmentative;
12/13	donnent aux substantifs une valeur diminutive, l'augmentatif et le diminutif pouvant, bien sûr, s'interpréter avec une nuance méliorative ou péjorative;
14	désigne des réalités abstraites ou le diminutif;
15	est consacrée - aux infinitifs substantivés,

⁷ Dans ce contexte, la notion de «catégorie sémantique» serait à rapprocher de celle de «classe d'objets» (cf. Gross et Clas 1997).

16 - aux différentes parties du corps (comme 5/6);
regroupe les diverses modalités de lieux.

Le sémantisme ci-dessus n'est cependant pas à prendre au sens absolu. En effet, il n'est pas rare de rencontrer ça et là, d'une part, des mots appartenant aux mêmes catégories sémantiques alors qu'ils portent des préfixes de classes différentes; tout comme, d'autre part, des mots de même préfixe de classes peuvent appartenir à des catégories sémantiques bien distinctes. Cependant, le caractère sémantique flou dont il est ici question ne concerne que certains préfixes kiruúndi; car il semblerait que les dix-neuf préfixes de classes du proto-bantu (Guthrie 1971 et Meeussen 1967) aient eu des valeurs sémantiques nettes, une netteté qui se retrouve en totalité ou en partie dans l'une ou l'autre langue bantu (Spitulnik 1987).

MÉMOIRE SÉMANTIQUE DES RADICAUX

Étant donné la multiplicité des valeurs sémantiques qui peuvent être générées par un même morphème radical, on peut soutenir, sans risque de se tromper, que la valeur sémantique du morphème radical est rarement monovalente. Bien souvent, que ce soit dans les langues bantu ou dans bien d'autres langues, le radical est sémantiquement polyvalent, ambigu, parfois même flou.

En effet, lorsqu'on observe en (8) les significations si différentes des unités lexicales formées à partir du radical -ntu, il devient extrêmement difficile de donner le sens exact du dit radical.

Par ailleurs, il n'est pas rare que deux mots de radicaux identiques expriment des idées contraires, tout simplement à cause de la valeur sémantique de leurs affixes (préfixes, infixes ou suffixes).

Tenant compte des faits évoqués ci-dessus, on peut donc affirmer que la mémoire sémantique d'un mot se modèle et se clarifie au fur et à mesure que les autres morphèmes lexicaux s'agglutinent au morphème radical. Seul le résultat final de cette agglutination permet de clarifier le sens du mot, qui reste, malgré tout, axé sur le sens du radical.

Malgré ses contours sémantiques flous, le morphème radical demeure le repère-clé dans la construction du sens des mots, principalement grâce à la force d'attraction que le radical exerce sur les autres morphèmes venant clarifier son sémantisme, mais également par son adaptabilité, lui permettant de se combiner avec divers affixes, en s'appropriant, en quelque sorte, leur valeur sémantique.

MÉMOIRE SÉMANTIQUE DES AUTRES MORPHÈMES

Contrairement à la charge sémantique de l'augment, du préfixe et du radical, celle des autres morphèmes lexicaux est assez nette, ce qui leur permet d'influencer de manière déterminante la mémoire sémantique des mots qu'ils constituent.

Voici, par exemple, la signification nettement exprimée par certains morphèmes du tableau 1 :

(9)

- les infixes :

-ha- :	lieu :	<i>abatáhahára (a-ba-ta-ha-har-a)</i>
-ii- :	réfléchi :	<i>umwíimenyerezo (u-mu-ii-meny-ir-ir-i-o)</i>
-ta- :	négation :	<i>abatáhahára (a-ba-ta-ha-har-a)</i>

- suffixes :

-an- :	ensemble :	<i>umutúuhúrana (u-mu-ta-uh-ur-an-a)</i>
-ir- :	applicatif :	<i>agatéerantúmba (a-ka-ta-ir-a-n-tiimb-a)</i>
-uur- :	reversif :	<i>imigondooro (i-mi-goond-uur-o)</i>

- finales :

-a- :	résultat :	<i>Umutúuhúrana (u-mu-ta-uh-ur-an-a)</i>
-e- :	état :	<i>urushíngé (u-ru-shiing-e)</i>
o- :	action/résultat :	<i>umwíimenyerezo (u-mu-ii-meny-ir-ir-i-o)</i>

Tous ces morphèmes servent de repères à la mémoire sémantique du mot et ont chacun une valeur sémantique propre, mais pas totale, car non autonome. Toutes ces valeurs sémantiques se prennent mutuellement pour référence et se repèrent principalement à la valeur sémantique du radical qu'elles viennent compléter.

2. AU-DELÀ DES SUBSTANTIFS

Les morphèmes lexicaux décrits dans les paragraphes précédents ne sont pas la propriété exclusive de la catégorie nominale. En effet, grâce aux nombreux mécanismes de création lexicale, et plus particulièrement la dérivation, plusieurs unités lexicales issues de catégories différentes partagent les mêmes morphèmes, et par conséquent, les mêmes repères mémoriels. Elles partageront, entre autres, le morphème radical qui, comme nous l'avons déjà signalé, sert à construire une quantité foisonnante de mots appartenant à des catégories lexicales variées.

En outre, même si les mots entrant dans la construction d'un énoncé possèdent, chacun, une mémoire particulière avec des repères propres, ils doivent tous partager des repères mémoriels communs, afin d'assurer une cohérence grammaticale de l'énoncé. C'est pour rester conforme à cette cohérence que, en kiruúndi, comme évoqué en 1.2.2, les mots prédictifs s'accordent en classe et en nombre avec le substantif argument de chaque énoncé minimal. Bien entendu, pour ce genre d'opérations le repère mémoriel le plus influent reste le préfixe de classe qui, dans les langues bantu, commande tous les accords. C'est ainsi que, pour sa charge grammaticale, le préfixe de classe est considéré comme repère mémoriel par excellence.

Les limites du présent travail ne nous permettent pas de rentrer dans les détails des fonctionnements morphologiques des différentes unités lexicales, car ils ressemblent, à quelques différences près, à ceux du substantif. Du reste, ce dernier n'a pas, lui non plus, été analysé à fond, à cause de ces contraintes liées à la nature même de ce travail. À la section précédente, nous avons voulu tout simplement montrer que les structures morphologiques et les repères mémoriels qu'ils représentent, au sein d'unités lexicales comme le verbe, l'adjectif, le pronom, etc. sont similaires à ceux du substantif. De plus, comme vu précédemment, toutes ces unités lexicales sont en constante interaction avec le substantif qui en commande les accords.

3. AU-DELÀ DU KIRUÚNDI

Tout comme nous n'avons pas voulu limiter nos analyses au seul substantif kiruúndi, nous voudrions, de la même manière, étendre leur portée, ne fût-ce que sommairement, à l'ensemble des langues bantu. En effet, la structure morphologique du nom kiruúndi semble correspondre au modèle morphologique type des noms en langues bantu; car, comme nous l'avons déjà signalé ça et là, lors de nos commentaires, l'essentiel des caractéristiques morphologiques du kiruúndi se retrouvent dans toutes les langues bantu.

Effectivement, mis à part l'augment absent du substantif dans l'une ou l'autre langue bantu, toutes partagent l'essentiel des morphèmes décrits en 1.

Du reste, si Bleek (1862) a donné le nom *bantu* à la famille linguistique dont il est ici question, c'est parce que les études historico-comparativistes ont montré que toutes les langues appartenant à cette famille utilisent le mot *muntu* (ou des formes phonétiques apparentées) pour nommer le concept de 'personne humaine', tandis que le nom *bantu* (ou des formes phonétiquement apparentées) désigne le même concept au pluriel; d'où les préfixes de classes 1 et 2 (-*mu*- et -*ba*-) réservés aux humains.

Les langues bantu disposent donc d'un héritage linguistique commun, aisément détectable à travers leurs comportements linguistiques, et plus particulièrement au niveau des structures morphologiques. Ce sont ces comportements linguistiques qui ont permis de reconstituer l'ancêtre commun de ces langues avec une relative facilité, et d'en comprendre les mécanismes de différenciation.

Par des distances linguistiques — liées, entre autres, aux mouvements des populations et à l'effet du temps — les membres de cette famille linguistique ont progressivement pris des identités linguistiques différentes, en modifiant petit à petit les structures phonologiques ainsi que les valeurs linguistiques de certains morphèmes.

Malgré leurs différences, les langues bantu ont cependant gardé un grand stock d'éléments linguistiques communs et des comportements linguistiques communs, notamment :

- la création lexicale régulière et prolifique, par agglutination autour du radical, la plupart des morphèmes radicaux étant communs à plusieurs langues;
- l'accord grammatical strict autour de la classe, le repère mémoriel classificatoire étant le préfixe nominal;
- ...

L'on comprend donc que les langues bantu avaient, au départ, les mêmes repères mémoriels, dont les plus importants étaient le radical et le préfixe de classe. Ces morphèmes sont toujours considérés comme les deux éléments lexicaux fondamentaux, car ils sont indispensables dans la construction de n'importe quelle unité lexicale en langue bantu.

C'est pour leur nature lexicale essentielle que nous avons pris le préfixes et le radical comme modèles illustratifs des repères mémoriels, le premier comme repère mémoriel par excellence pour sa charge grammaticale, et le second comme repère mémoriel d'importance capitale au niveau sémantique.

4. AU-DELÀ DE LA MÉMOIRE LINGUISTIQUE

Nous avons essayé de montrer, ci-dessus, comment les langues bantu sont organisées en de riches et réguliers réseaux de morphèmes lexicaux, qui constituent des repères incontournables de la mémoire linguistique des énoncés et de leurs constituants.

Après avoir observé des réseaux si exubérants et en même temps si bien organisés, nous croyons que les éléments de ces structures morphologiques gagneraient à être répertoriés et classés dans des dictionnaires, tout comme on le fait à grande échelle pour les unités lexicales, et dans une moindre mesure pour les locutions.

Bien entendu, pour être optimalement exploitables, ces structures seraient répertoriées dans des dictionnaires électroniques, afin de permettre un accès rapide à l'information et un inventaire le plus exhaustif de données, sans beaucoup de contraintes spatiales.

De plus, pareil dictionnaire permettrait que les repères morphématiques dépassent le niveau de l'habituelle mémoire linguistique — individuelle et collective — exclusivement réservée à ceux qui maîtrisent la langue, pour atteindre la mémoire numérique, directement accessible par la machine, et par conséquent, théoriquement exploitable par tous, y compris les non-locuteurs de la langue concernée.

Pour une meilleure efficacité, le dictionnaire dont il est question ici serait structuré comme une base de données contenant les listes des différents morphèmes⁸ avec leurs valeurs linguistiques respectives, comme la position du morphème dans le mot, ses cooccurents, sa valeur sémantique, etc. Toutes ces informations seraient progressivement rentrées dans la base de données au fur et à mesure de la décomposition morphologique de mots, tirés de textes expressément choisis à cet effet.

Nous avons vu, par ailleurs, comment les morphèmes du kiruúndi sont organisés en réseaux de repères mémoriels. Ces réseaux pourraient aisément servir de guides pour créer des liens dans la base de données. Celle-ci serait, par conséquent, une base de données relationnelle, susceptible de permettre une reconstitution automatique des liens intra-lexicaux unissant les différents morphèmes d'un mot, ainsi que les liens extra-lexicaux qui commandent les accords de classes dans un énoncé.

Une base de données pareille prendrait sûrement beaucoup de temps et d'efforts avant d'être exploitable. En retour, eu égard à la représentativité, à la richesse et à la variété de ses enregistrements, la base de données serait d'un intérêt linguistique certain, primordialement en linguistique théorique, à cause des nombreuses analyses statistiques qu'elle permettrait de réaliser, comme par exemple :

- établir la fréquence d'utilisation de tel ou tel morphème lexical;

⁸ Comme les radicaux forment des listes ouvertes, la base de données enregistrerait les morphèmes radicaux au fur et à mesure qu'ils seraient identifiés.

- résoudre la question non encore bien tranchée (Hyman 1993; Meeussen 1967 etc.) au sujet de la compatibilité et de l'ordre des affixes⁹.

Bien entendu, pour une plus grande efficacité, la base de données serait couplée à une interface d'analyse et de synthèse des unités lexicales, qui comprendrait, entre autres :

- un programme intégrant les règles phonologiques régissant les combinaisons des morphèmes, ainsi que les règles d'accords et de concordances;
- un programme de découpage morphologique, capable de décomposer chaque mot en ses morphèmes. Ce programme fonctionnerait comme les programmes de parsing qui décomposent les propositions en vue d'en dégager les unités lexicales constitutives;
- d'autres informations linguistiques pertinentes, au sujet des mécanismes de découpage ou de combinaisons morphologiques;
- ...

Évidemment, cette interface analytico-synthétique aurait comme dictionnaire-repère la base de données, avec ses unités lexicales et morphologiques. Ceci signifie que l'interface se servirait des mots et des morphèmes de la base de données comme modèles d'identification d'autres mots dans les textes à analyser ou pour vérifier l'exactitude des mots et morphèmes reconstitués.

Ainsi, du point de vue analytique, et à l'aide de l'interface ci-dessus, cette base de données permettrait d'analyser des textes, pour y détecter les unités lexicales et en extraire les morphèmes lexicaux. Ces derniers pourraient être réintroduits dans la base de données et servir ainsi à l'enrichir.

Par ailleurs, comme la base de données contiendrait les valeurs sémantiques de ses morphèmes, l'étape précédente d'extraction des morphèmes pourrait être complétée par leur interprétation sémantique, et contribuer ainsi à la compréhension de mots analysés, éventuellement à leur traduction, peut-être même à la traduction automatique de certains textes.

Sur le plan de la synthèse automatique, et par le truchement de l'interface décrite précédemment, la base de données serait également la source tout indiquée pour la génération automatiquement des mots, comme ceux naturellement produits par les locuteurs de langues bantu, par combinaison des différentes valeurs sémantiques des morphèmes. Ces mécanismes de création synthétique des mots seraient de précieux adjuvants accélérant la création de nouveaux termes, principalement pour désigner les concepts techniques nouveaux¹⁰. Cependant, personne ne devrait en exagérer l'utilisation, sans risque de retirer aux mots leur essence naturelle. La génération de mots de synthèse ne serait qu'une solution de dernier recours, utilisable seulement si la mémoire linguistique naturelle est à cours

⁹ On sait, jusqu'à présent, que plusieurs infixes et suffixes peuvent se retrouver dans une même unité lexicale, mais ils ne se combinent pas de manière anarchique, tout comme ils n'acceptent pas n'importe quel morphème. À l'heure actuelle, il ne semble pas y avoir, en ce qui concerne le kiruúndi en tout cas, d'étude pointue sur ces règles — que tout locuteur suit sans s'en rendre compte — régissant l'ordre et la combinabilité de ces affixes.

¹⁰ Nous avons déjà expérimenté pareilles procédures en générant automatiquement un vocabulaire kiruúndi de termes de cuisine ainsi que leurs traductions françaises.

d'inspiration; et le produit de cette génération devraient impérativement être soumis à la vérification et à l'approbation d'un nombre significatif de locuteurs usagers avant d'être soumis au grand public.

CONCLUSION

Nous sommes parti d'une analyse du substantif kiruúndi, pour illustrer la régularité des structures morphologiques de l'ensemble du lexique kiruúndi. Nous avons montré comment ces structures fournissent à la langue une mémoire lexicale très prolifique et en même temps rigoureusement ordonnée par les morphèmes lexicaux. C'est pourquoi les morphèmes ont été décrits comme des repères indispensables dans l'élaboration de la mémoire lexicale, une mémoire agissante à tous les niveaux linguistiques.

Comme la langue kiruúndi a beaucoup d'affinités avec les autres langues bantu, il n'a pas été difficile de justifier que la mémoire lexicale du kiruúndi est similaire à celles des autres langues-sœurs.

Nous avons ensuite parlé des possibilités de classer les éléments de cette mémoire linguistique, à la fois si abondante et si régulière, en stockant dans une base de données les différents morphèmes, ainsi que leurs valeurs linguistiques respectives.

Une base de données relationnelle comme celle décrite ci-dessus est susceptible de nombreuses applications, en linguistique comme dans d'autres disciplines connexes. Nous avons sommairement évoqué certaines des applications que cette base de données pourrait permettre de réaliser, que ce soit en linguistique théorique ou en linguistique appliquée à l'analyse ou à la synthèse automatique des unités lexicales.

Dans les conditions optimales, notre modèle de base de données pourrait devenir contrastif et contenir des informations linguistiques en provenance de plusieurs langues bantu, afin d'offrir une vue plus large de leurs repères mémoriels lexicaux. Ainsi, la valeur linguistique de chaque morphème dans une langue serait mise en parallèle avec la valeur linguistique du même morphème dans les autres langues bantu. Conséquemment, les valeurs linguistiques identiques dans différentes langues seraient également mises en relation avec les morphèmes correspondants.

Pour donner une vue d'ensemble des repères mémoriels lexicaux en langues bantu, l'aspect contrastif porterait sur le maximum possible de langues, et idéalement sur toutes. Cependant, ce projet paraît trop gros pour pouvoir être entrepris dans l'immédiat.

Malgré tout, si le projet de base de données se montre concluant dans l'une ou l'autre langue, il sera alors possible de l'étendre, petit à petit, sur d'autres langues¹¹ et d'entreprendre une mise en relation de leurs bases de données, créant ainsi une méga base de données multilingues. Seul l'avenir nous dira si le rêve que nous caressons depuis longtemps, de créer une base de données contrastives sur l'ensemble des données morpho-lexicales des langues bantu, serait réalisable ou non.

¹¹ Toutes les langues devront être, bien entendu, préalablement soumises au même modèle de description linguistique.

RÉFÉRENCES

- BASTIN, Y. (1985) : *Les relations sémantiques dans les langues bantoues*, Bruxelles, ARSOM, 86 p.
- BENDOR-SAMUEL, J. & R. L. HARTELL (1984) : *The Niger-Congo Languages : a classification and description of Africa's largest language family*, New York, University Press of America, 505 p.
- BLEEK, W. H. I. (1862) : *A comparative grammar of South African Languages*, part 2, Cape Town, Juta.
- FAÏK-NZUJI, C. M. (1992) : *Éléments de phonologie et de morphologie des langues bantu*, Louvain-la-Neuve, Peeters, 163 p.
- FODOR, I. (1982) : *A fallacy of contemporary linguistics. J. H. Greenberg's classification of the african languages and his «comparative method»*, Hamburg, Buske.
- GREENBERG, J. H. (1970) : *The languages of Africa*, 3rded., Bloomington, Indiana University, 180 p.
- GROSS, G. et A. CLAS (1997) : «Synonymie, polysémie et classes d'objets», *Meta*, vol 42, n° 1, (mars 1997), Montréal, PUM, pp. 147-154.
- GUTHRIE, M. (1971) : *Comparative bantu*, Farnborough, Gregg International Publishers, vol. 1-143 p., vol. 2-180 p., vol. 3-326 p., vol. 4-248 p.
- HYMAN, Larry M. (1993) : «Conceptual issues in the comparative study of the bantu verb stem», Mufwene S. S. & Moshi L., *Topics in African Linguistics*, Philadelphia, John Benjamins Publishing Company, pp. 3-34.
- LEMARÉCHAL, Alain (1989) : *Les parties du discours. Sémantique et syntaxe*, Paris, PUF, 272 p.
- MANN, M. & D. DALBY (1987) : *A thesaurus of african languages : a classified and annotated inventory of the spoken languages of Africa with an appendix on their written representation*, London, Hans Zell Publishers, 325 p.
- MEEUSSEN, A. E. (1959) : *Essai de grammaire rundi*, Tervuren, AMRAC, 236 p.
- MEEUSSEN, A. E. (1967) : *Bantu grammatical reconstructions*, Tervuren, AMRAC, 236 p.
- MEINHOF, C. (1899) : *Grundriss einer Lautlehre der Bantusprachen nebst einer Anleitung zur Aufnahme von Bantusprachen*, Berlin, Dietrich Reimer.
- MONINO, Y. (1995) : *Le proto-gbaya : Essai de linguistique comparative sur vingt-et-une langues d'Afrique centrale*, Paris, Peeters, 725 p.
- MUFWENE S. S. & L. MOSHI (1993) : *Topics in African Linguistics*, Philadelphia, John Benjamins Publishing Company, 307 p.
- NTAKIRUTIMANA, J. B. (1993) : *La créativité lexicale en kiruúndi, Étude lexico-sémantique de la néologie dans les manuels de lecture à l'école primaire*, mémoire de Maîtrise, Montréal, Université de Montréal, 119 p.
- NTAKIRUTIMANA, J. B. (1993) : *Vers une traduction automatique kiruúndi-français : Approche présémantique des substantifs de la terminologie culinaire*, Séminaire de Traduction automatique, hiver 1993, travail de fin de session, Montréal, Université de Montréal, (inédit).
- RODEGEM, F. M. (1970) : *Dictionnaire rundi-français*, Tervuren, AMRAC, 644 p.
- SPITULNIK, D. A. (1987) : *Semantic Superstructuring and infrastructuring : Nominal classes struggle in Chibemba*, Bloomington, Indiana University Linguistic Club, 146 p.

LES CANADIANISMES EN FRANÇAIS ET EN ANGLAIS¹

Roda P. ROBERTS et Chantale GRENON-NYENHUIS

Université d'Ottawa, Canada

INTRODUCTION

Comme le signalent Jean Dubois et Claude Dubois (1971 : 8), les dictionnaires sont «des objets culturels, intégrés comme tels à cette culture : ils témoignent d'une civilisation». Cet avis, partagé par beaucoup, est fondé sur le double constat que le dictionnaire a pour vocation première d'informer sur la langue, et que la langue et la culture sont étroitement liées (Rialland-Addach 1995 : 91). En effet, le dictionnaire s'articule sur deux plans en termes hjelmsléviens : celui de l'expression, «une langue et son lexique, des usages et leurs vocabulaires, certains discours et leurs occurrences», et celui du contenu, «un univers exprimé, des visions culturelles du monde, des articulations conceptuelles correspondant à des classes extensionnelles» (Rey et Delasalle 1979 : 5).

Mais si le dictionnaire unilingue est bien reconnu comme «*le livre privilégié de référence* à la connaissance et au savoir linguistique et culturel» d'une communauté (Dubois et Dubois 1971 : 8), le dictionnaire bilingue l'est beaucoup moins. Selon Dubois et Dubois (1971 : 34-35), le dictionnaire bilingue repose sur l'identité supposée des concepts et des manières de percevoir le monde dans deux langues; ainsi, «les cultures transmises peuvent être ramenées à un certain nombre d'invariants communs : les visions du monde sont identiques, imposées par les principes logiques qui sont à la base du fonctionnement d'une pensée universelle». En d'autres termes, selon eux, le dictionnaire bilingue présente la culture partagée plutôt que les particularités culturelles de chaque langue, car il cherche à donner des paires de synonymes. Cependant, on commence peu à peu à reconnaître qu'on ne peut s'attendre à trouver des «équivalents complets» dans le dictionnaire bilingue, car «*the conceptual world evolves differently in different languages as a result of, for example, historical, geographical, social, cultural, and economic differences between the countries where the different languages are used*» (Svensén 1993 : 140). On commence à accepter l'idée que le dictionnaire bilingue peut inclure des équivalents partiels ou même aucun équivalent, pour autant que le lexicographe identifie et indique les différences entre les deux langues (Svensén 1993 : 142). Cette nouvelle façon de voir le rôle du dictionnaire bilingue en fait, lui aussi, un ouvrage culturel, qui reflète non pas la culture d'une communauté linguistique, ni même la culture partagée, mais deux cultures différentes.

¹ Cette recherche a été subventionnée par le Conseil de recherches en sciences humaines du Canada.

Un tel ouvrage est d'autant plus nécessaire quand deux langues et donc deux cultures coexistent, comme c'est le cas au Canada avec l'anglais et le français. Même si ces deux langues sont utilisées sur le même territoire, la culture véhiculée par chacune n'est pas la même. En outre, ce qui complique d'autant plus la production d'un dictionnaire canadien bilingue, c'est que l'anglais et le français du Canada diffèrent sur certains points du français de France et de l'anglais britannique et américain, car ces deux langues ont connu une évolution différente dans ce pays et reflètent des réalités différentes. En d'autres mots, l'anglais et le français canadiens ont des particularités linguistiques et culturelles non seulement l'un par rapport à l'autre, mais aussi par rapport à l'anglais et au français tels qu'ils sont utilisés ailleurs. Ces particularités linguistiques canadiennes, souvent appelées «canadianismes», se manifestent dans la prononciation, la syntaxe, la morphologie, mais surtout le vocabulaire. Ce sont les canadianismes lexicaux qui seront examinés ici dans l'optique de la production d'un nouveau dictionnaire bilingue, le *Dictionnaire bilingue canadien* (DBC), qui, ses rédacteurs l'espèrent tout au moins, sera l'ouvrage culturel par excellence de ce pays bilingue².

1. L'ORIGINE DES CANADIANISMES

On peut attribuer l'origine des canadianismes en anglais ainsi qu'en français aux circonstances dans lesquelles ces langues se sont implantées au Canada. Le fait que l'historique du français au Canada diffère de celui de l'anglais au Canada explique en grande partie les divergences entre les canadianismes dans ces deux langues.

1.1 Historique du français canadien

Le Canada est découvert en 1534 par Jacques Cartier, qui prend possession du territoire au nom du roi de France. Le français est donc la première langue européenne qui tente de conquérir ce vaste territoire. Mais les ambitions françaises se limitant aux richesses qui peuvent être acquises dans ce pays, la Nouvelle-France n'est pas immédiatement colonisée. Ce n'est qu'après la fondation de la ville de Québec par Samuel de Champlain en 1608 et la fondation subséquente de Ville-Marie et de Trois-Rivières que la population francophone commence à s'accroître, quoique à un rythme relativement lent. Nous parlons de la population francophone de l'époque, même s'il est difficile de déterminer si le français s'est développé avant ou après l'arrivée des immigrants venus de France. Selon l'hypothèse traditionnelle, les patois, originaires surtout de l'Île-de-France, de l'Ouest et du Nord-Ouest, se sont rencontrés en arrivant en Nouvelle-France et ont constitué la base de ce qui est devenu le français canadien. Cependant, d'après une nouvelle théorie (Mougeon et Beniak 1994 : 1-55), il est probable que les immigrants, qui devaient passer un certain temps dans la même ville avant de s'embarquer pour l'Amérique, aient appris le français comme langue commune avant d'entreprendre le voyage.

Quoi qu'il en soit, le français au Canada est influencé dès ses débuts par les langues autochtones, dont il emprunte certains vocables pour désigner de nouvelles réalités auxquelles sont exposés les Européens lors de leur arrivée sur le continent nord-américain

² Il s'agit d'un projet interuniversitaire, mené en collaboration par l'Université d'Ottawa, l'Université de Montréal et l'Université Laval, qui vise la production en 2003 d'un dictionnaire anglais-français, français-anglais reflétant les usages canadiens.

(ex. **babiche**, **boucane**, et **toboggan**, issus de langues amérindiennes, et **mukluk** et **umiak**, relevés des dialectes inuit).

Mais l'impact critique sur le français canadien vient de l'anglais, qui commence à s'établir en Amérique du Nord à partir de 1713, quand les Britanniques s'installent sur la côte est de ce qui constitue maintenant les États-Unis, dans le but de coloniser ce nouveau territoire. En peu de temps, la population de langue anglaise devient supérieure à celle de langue française. Les Britanniques commencent à revendiquer une partie des terres canadiennes et, en 1759, s'emparent de la ville de Québec. Enfin, en 1763, le Traité de Paris oblige la France à céder la Nouvelle-France aux Anglais et le français du Canada est isolé du français de France. C'est à cette époque que l'anglais est introduit dans la colonie et qu'une certaine assimilation des francophones et de leur culture commence.

Si l'Acte de Québec, signé en 1774, permet aux francophones de revenir au système de droit civil français et d'occuper des postes politiques tout en continuant de pratiquer la religion catholique, le rapport de Lord Durham, rédigé à la suite des Rébellions des Patriotes en 1837-1838, préconise l'assimilation des francophones (Mallory 1984 : 340) et recommande l'union du Bas-Canada (majoritairement francophone) et du Haut-Canada (majoritairement anglophone). En 1840, l'Acte d'Union entre en vigueur et la politique d'assimilation de la population francophone se poursuit.

Mais l'Église catholique, qui exerce une grande influence sur les francophones du Bas-Canada, s'assure de garder séparées la population francophone et la population anglophone. Ainsi, le français survit, du moins au Bas-Canada, même si les liens avec le français de France sont rompus et si le français occupe une position d'infériorité par rapport à l'anglais. Mais le français évolue différemment sur les deux continents; en Amérique du Nord, son évolution est influencée par l'anglais. Il faut attendre la Révolution tranquille des années 60 pour qu'un effort concerté soit fait afin de libérer le français de l'anglais et pour améliorer la qualité du français canadien. Grâce aux travaux terminologiques de l'Office de la langue française, créé en 1961, et aux lois linguistiques approuvées par le Gouvernement du Québec, le français reprend sa vigueur et devient un outil de travail quotidien.

Ce bref résumé historique explique d'abord pourquoi il y a des canadianismes en français et donne ensuite une indication des types de canadianismes qui s'y retrouvent : emprunts aux langues autochtones, mots et sens tombés en désuétude en France, anglicismes, et terminologie moderne particulière.

1.2 Historique de l'anglais canadien

Comme nous l'avons signalé précédemment, les Britanniques s'installent en Amérique du Nord après les Français. Comme ces derniers, ils empruntent des mots autochtones pour décrire les réalités nord-américaines; ces emprunts s'effectuent généralement directement aux langues autochtones, mais parfois aussi à la suite des contacts avec la population francophone qui avait déjà fait des emprunts à ces langues.

Devenant rapidement supérieurs en nombre, les Anglais réclament le territoire français au Canada et l'acquièrent après le Traité de Paris en 1763. À partir de cette date, l'anglais établit sa dominance au Canada.

Mais de quel anglais s'agit-il ? Les anglophones qui viennent s'installer au Canada ne sont pas uniquement originaires de Grande-Bretagne. En fait, beaucoup d'Américains immigrèrent vers le nord, surtout après la Déclaration d'indépendance des États-Unis en 1776, quand un bon nombre de Tories et de Loyalistes sont obligés de s'exiler à cause de leurs convictions politiques. Ainsi se développe petit à petit un anglais canadien qui n'est ni britannique ni américain, mais qui possède certaines caractéristiques des deux langues, car les anglophones du Canada, contrairement aux francophones, ont toujours gardé contact avec la Grande-Bretagne et les États-Unis. En outre, bien entendu, l'anglais canadien a acquis un certain nombre de mots pour désigner des réalités typiquement canadiennes. Mais la grande ressemblance de l'anglais canadien avec l'anglais américain, d'une part, et l'anglais britannique, de l'autre, a longtemps entravé la reconnaissance des particularités de cette variété de langue.

Ce bref historique de l'anglais canadien laisse supposer que la plupart des canadianismes dans cette langue seraient des mots autochtones et des mots désignant les réalités propres au Canada.

2. IDENTIFICATION DES CANADIANISMES

Ce sont ces canadianismes en anglais et en français qui reflètent la culture typiquement canadienne. Il est donc de rigueur de les inclure dans un dictionnaire canadien bilingue. Mais comment les identifier sans effectuer une comparaison systématique, qui serait longue et pénible, de tout le vocabulaire de l'anglais canadien avec les vocabulaires américain et britannique et une comparaison du vocabulaire du français canadien avec celui du français hexagonal ? Le DBC a utilisé quatre types de sources différentes pour y arriver.

Le point de départ était la documentation lexicographique. D'abord, nous avons examiné les dictionnaires canadiens anglais et français pour constituer des listes préliminaires. Nous nous sommes servis principalement du *Gage Canadian Dictionary (GAGE)* (1983), du *Funk and Wagnalls Canadian College Dictionary (FUN)* (1989) et du *Penguin Canadian Dictionary (PEN)* (1990) pour l'anglais, ainsi que du *Dictionnaire du français Plus (PLUS)* (1988) et du *Dictionnaire québécois d'aujourd'hui (RQ)* (1992) pour le français. En dépouillant les dictionnaires canadiens anglais, nous avons repéré tous les mots et les emplois marqués «Canadian». Et quand ces ouvrages n'étaient pas d'accord sur la marque géographique (c'est-à-dire si le mot ou l'emploi n'était pas marqué dans un des dictionnaires ou portait une autre marque géographique comme «North American»), nous avons fait des vérifications supplémentaires dans un dictionnaire américain, le *Random House Webster's College Dictionary (RHWEB)* (1991), ainsi que dans un dictionnaire britannique, le *Collins English Dictionary (COLL)* (1986). Le processus d'identification des canadianismes français dans les dictionnaires n'était pas identique, pour la simple raison que les dictionnaires canadiens mentionnés ci-dessus ont comme politique de ne pas marquer ce qui est canadien, car ils prennent comme point de référence le français du Canada. Ainsi, il a fallu laisser le repérage des canadianismes français à nos spécialistes du français canadien, qui les identifiaient d'abord en jumelant l'examen des deux dictionnaires canadiens et leurs propres connaissances linguistiques, puis en cherchant les éléments retenus dans les dictionnaires hexagonaux (le *Petit Robert*, le *Lexis* et le *Petit Larousse*) : l'absence de ces éléments dans ceux-ci ou leur marquage comme mot «canadien» permettait de confirmer leur statut de canadianisme. Les listes préliminaires des canadianismes anglais et français, dressées comme nous l'avons indiqué ci-dessus étaient ensuite élargies en

consultant d'autres ouvrages lexicographiques plus restreints et plus spécifiques, comme le *Concise Dictionary of Canadianisms (CDC)* (1973), *Richesses et particularités de la langue écrite au Québec (RP)* (1980), et le *Dictionnaire du français canadien (DCF)* (1990), ainsi que le fichier lexical du Trésor de la langue française au Québec (TLFQ)³.

Des consultants et des informateurs constituaient notre deuxième source d'information. Nous avons déjà signalé le rôle essentiel que jouaient nos spécialistes du français canadien dans l'établissement de la liste initiale. Ceux-ci, à leur tour, ont consulté au besoin les lexicographes du TLFQ. Pour l'anglais, nous avons tenté de faire une consultation plus large en faisant appel aux associations d'études canadiennes, sans grand succès, cependant.

Comme troisième source d'information sur les canadianismes, nous avons utilisé des articles traitant de l'anglais et du français canadiens. Beaucoup d'entre eux sont des articles de recherche (ex. R.J. Gregg 1989 : 151-187; C. Poirier 1980 : 43-89). D'autres sont des articles journalistiques. L'examen de ces articles nous a permis de nous assurer que nous avons relevé tous les canadianismes importants.

Enfin, la quatrième source d'information, peut-être la plus innovatrice, était, et continue d'être, des corpus de textes. Le DBC se sert principalement de deux bases de données textuelles : TEXTUM et TRANSBASE. TEXTUM, qui a été établie en fonction des besoins du DBC, contient des textes surtout journalistiques, mais aussi littéraires et scientifiques; tandis que la plupart des textes sont canadiens d'origine, cette base de données inclut un certain nombre de textes du français hexagonal et de l'anglais américain, qui nous permettent de faire des comparaisons afin de mieux identifier les usages canadiens⁴. TRANSBASE, base de données constituée par le Centre d'innovation en technologies de l'information, contient un corpus de traduction qui comporte des textes de départ et des traductions couvrant une période de huit ans du journal des débats du Parlement canadien, communément appelé le *Hansard*.

En utilisant ces quatre sources, nous avons établi deux listes de canadianismes (l'une pour l'anglais, l'autre pour le français) à inclure dans le DBC. Il faut noter qu'il s'agit de listes ouvertes : d'abord parce que, étant donné la durée du projet (une dizaine d'années), de nouveaux canadianismes risquent d'apparaître, surtout en français où la terminologie technique est encore en train de s'établir; puis, parce que nous arrivons parfois à repérer des sens canadiens «oubliés» lors de l'étude d'une concordance pour un mot-vedette particulier. Ce n'est, en effet, qu'en examinant TEXTUM pendant la préparation de l'article pour le nom *aîné* que nous avons identifié l'un de ses sens, celui d'une personne d'une tribu autochtone, qui est généralement âgée et qui a beaucoup d'influence sur la vie de la tribu. C'est un des sous-corpus canadiens, constitué d'articles de la *Presse canadienne française*, qui a révélé ce sens qui ne se trouve dans aucun dictionnaire, même canadien. Voici quelques exemples d'*aîné* dans ce sens, tirés du sous-corpus :

³ Nous tenons à exprimer notre reconnaissance envers l'équipe du TLFQ pour toute l'aide qu'elle nous a fournie.

⁴ Bien que TEXTUM ne contienne pas de textes britanniques, nous nous servons des *MicroConcord Corpus Collections* sur CD-ROM, qui contiennent d'une part des textes journalistiques, d'autre part des textes «académiques» dans une variété de domaines.

Presse canadienne française

>> aine

1 : 2726 matches

>> "aine "

2 : 1098 matches

>> 2 near "cri "

3 : 3 matches

>> pr

179365963,.. demandé à un aîné cri, à l'occasion du procès qui a précédé le..

176165820,.. demandé à un aîné cri, à l'occasion du procès qui a précédé le..

164006111,.. sagesse de l'aîné Cri que vous citez, en disant modestement com..

>> pr.200 shift.-100

164006011,..n souci d'équité et de justice. En terminant, permettez-nous de vous assurer que nous respectons la sagesse de l'aîné Cri que vous citez, en disant modestement comme lui que nous ne possédons pas la..

179365863,..u Grand conseil des Cris du Québec et vice-président de l'Administration régionale crie. Quand on a demandé à un aîné cri, à l'occasion du procès qui a précédé le jugement Malouf dans l'affaire de la..

>> 2 near inuit

10 : one match

>> pr

293531366,..lo Pudlat, un aîné respecté parmi les artistes inuit du Canada j..

>> pr.200 shift.-100

293531266,..anisée dans l'Arctique canadien. L'exposition Pudlo. Une célébration est consacrée à l'oeuvre de Pudlo Pudlat, un aîné respecté parmi les artistes inuit du Canada jusqu'à sa mort récente. L'ouverture..

Pour nous assurer qu'il s'agissait bien d'un canadianisme, nous avons fait la même recherche dans le sous-corpus hexagonal, *Le Monde*. Comme le montre l'historique de nos recherches, présenté ci-dessous, ce sens ne semble pas exister en France.

Le Monde

>> aine

1 : 315 matches

>> "aine "

2 : 146 matches

>> 2 near cri

3 : one match (not relevant)

>> pr

46430207,..Paris, Dorbon aîné. DOC:AVEC UN DESSIN TIRE D'UNE BANDE DESSIN..

3. SÉLECTION DES CANADIANISMES POUR LA NOMENCLATURE DU DBC

En utilisant les procédures décrites ci-dessus, les chercheurs du DBC ont identifié un grand nombre de canadianismes en français et en anglais à inclure dans le dictionnaire. Mais nos listes ne sont pas exhaustives car, pour avoir droit de cité dans notre dictionnaire, les canadianismes doivent répondre à certains critères, qui ne sont pas exactement les mêmes pour l'anglais et le français.

Pour l'anglais, nous avons retenu presque tous les canadianismes qui se trouvaient dans deux sources lexicographiques (autres que les dictionnaires historiques) ou dans les corpus, car il ne s'agit pas de plusieurs milliers de mots ou d'emplois : nous n'avons éliminé que ceux dont le statut de canadianisme n'est pas confirmé, ceux qui sont utilisés seulement dans une région particulière (ex. le mot terre-neuvien **airsome**, adjectif, dans le sens de froid ou frais) ainsi qu'un certain nombre de canadianismes qui seraient aujourd'hui considérés comme étant archaïques (ex. **dime party**, dans le sens d'une soirée à laquelle on pouvait assister en offrant un don de dix sous) ou qui reflètent une réalité qui ne serait plus comprise aujourd'hui (ex. **fur board**, dans le sens de «thin board inserted in a cased skin to hold its shape until dry» - CDC). En appliquant ces quelques critères, nous nous sommes retrouvés avec moins de trois cents canadianismes lexicaux.

Pour le français, les critères de sélection ont été à la fois plus sévères et plus souples. Les canadianismes qui ont été exclus systématiquement, comme pour l'anglais, sont ceux qui sont régionaux, archaïques, et «historiques». Mais on a remplacé le critère de présence des canadianismes dans deux sources lexicographiques (non historiques) ou dans les corpus par celui de la fréquence relative de leur usage, selon le jugement de nos spécialistes du français canadien. Ce changement s'explique d'abord par le fait que les dictionnaires vraiment canadiens sont peu nombreux en français et que ni le *PLUS* ni le *RQ* ne sont considérés un reflet fidèle du français canadien, puis par le fait que beaucoup de canadianismes français, étant de niveau familier, ne se retrouvent pas dans les textes écrits qui constituent la plus grande partie de nos corpus. L'application de ces critères a mené à la sélection de quelques 3 000 canadianismes.

4. PROPORTION DE CANADIANISMES EN ANGLAIS ET EN FRANÇAIS

De prime abord, le nombre de canadianismes sélectionnés pour le DBC en anglais et en français peut sembler déséquilibré. En effet, le DBC comptera dix fois plus de canadianismes français que de canadianismes anglais. Mais un point important qui ressort de notre identification des canadianismes est qu'il existe beaucoup plus de canadianismes en français qu'en anglais.

Pour donner une idée plus claire de la proportion des canadianismes dans chaque langue (par rapport aux mots et aux emplois partagés avec d'autres variétés géographiques de chaque langue), examinons la nomenclature du *GAGE* et du *RQ*. Ces deux dictionnaires canadiens unilingues, dont nous nous sommes servis, entre autres, pour établir la nomenclature du DBC, incluent un plus grand nombre de canadianismes que les autres dictionnaires généraux canadiens. Même s'ils ne sont pas de taille comparable, ils se prêtent bien à une comparaison de la proportion des canadianismes dans chaque langue.

Ainsi, sur environ 70 000 entrées, le *GAGE* compte environ 800 canadianismes, si l'on inclut dans le total non seulement les mots ou les emplois qui contiennent la marque géographique *Cdn*, mais aussi ceux dont la particularité canadienne ne se manifeste que dans la définition; c'est donc dire que les canadianismes constituent moins de 1,5 % du total. Par contre, sur environ 40 000 entrées, le *RQ* compte à peu près 4 000 canadianismes; en d'autres termes, les canadianismes constituent 10 % du total. Ces statistiques très rudimentaires ont besoin d'être nuancées, car, d'une part, nous avons fait un calcul approximatif du nombre d'entrées et, d'autre part, les canadianismes ne sont pas tous présentés sous forme d'entrées séparées, parce qu'ils comprennent les emplois ainsi que les mots. Par ailleurs, bien que les deux dictionnaires se vantent de présenter des canadianismes, leurs préfaces sont muettes au sujet des critères utilisés pour les sélectionner. Mais, malgré ces réserves, on peut conclure que la proportion des canadianismes est beaucoup plus élevée en français qu'en anglais.

Le grand nombre de canadianismes français peut s'expliquer par les deux siècles d'isolement qu'a connu le Canada français. Cette période d'isolement a forcé le français canadien à se développer indépendamment du français hexagonal. Pendant ces deux cents ans, les francophones du Canada ont dû adapter leur langue aux réalités qui les entouraient. C'est dans le but de désigner certaines de ces réalités qu'ils ont donné des sens nouveaux à des mots qui étaient connus en France et qu'ils ont emprunté d'autres mots aux langues avec lesquelles ils étaient en contact. Cet isolement explique aussi pourquoi certains vocables ou sens qui sont sortis de l'usage ailleurs dans la francophonie sont encore employés régulièrement au Canada, s'ajoutant ainsi au stock de canadianismes.

D'autre part, le nombre apparemment moins élevé de particularités lexicales en anglais canadien pourrait être expliqué, nous croyons, par le fait que l'anglais canadien n'a jamais été isolé des autres sources d'anglais. Il a toujours joui de l'influence de l'anglais britannique et de l'anglais américain, même si l'importance de ces influences varie d'une époque à l'autre.

Ainsi, ce sont des facteurs historiques, politiques et même économiques qui semblent être responsables de l'important écart qui existe entre le nombre de canadianismes français et de canadianismes anglais.

5. TYPES D'UNITÉS LEXICALES QUE REPRÉSENTENT LES CANADIANISMES EN FRANÇAIS ET EN ANGLAIS

Il existe aussi des divergences entre les types d'unités lexicales que représentent les canadianismes en français et en anglais. En préparant la nomenclature du DBC, nous avons eu l'impression que la plupart des canadianismes en anglais sont des mots simples ou des mots composés, tandis que les canadianismes en français comprennent aussi des unités plus complexes comme les collocations et les expressions figées⁵. Pour confirmer cette impression, nous avons examiné à fond les canadianismes classés sous les lettres *D*, *F*, et

⁵ Dans cette partie de l'étude, nous ne faisons pas de distinctions entre les canadianismes de forme et les canadianismes sémantiques. Ainsi, si un sens canadien s'ajoute à un mot simple français commun au Canada français et à la France, on considère ici le mot simple comme canadianisme. La même chose s'applique aux mots composés, aux collocations et aux expressions figées.

S de la nomenclatures des canadianismes élaborée pour le DBC, les classant dans l'une ou l'autre de ces catégories⁶.

Des 65 canadianismes anglais à l'étude, 29 (soit 44,6 %) sont des mots simples (ex. **decker**, **dude**, **sourdough**); 35 (soit 53,8 %) sont des mots composés (ex. **deputy minister**, **diamond willow**, **fire hall**, **snow goose** et **sugaring off**); nous n'avons trouvé aucune expression figée, bien que nous ayons identifié une collocation (soit 1,5 %) (**to make a décharge**).

Par ailleurs, des 896 canadianismes français étudiés, 436 (soit 48,6 %) sont des mots simples (ex. **dépisteur/dépisteuse**, **désencrage**, **fichument**, **flasher**, **francophonisation**, **sacoche**, **salaison** et **siffleux**); 199 (soit 22,2 %) sont des mots composés (ex. **chat sauvage**, **Département de santé communautaire**, **directeur-gérant** (dans le domaine du sport), **disco-mobile**, **fonds de pension**, **sac d'école** et **sous-verre**); il y a 51 collocations (soit 5,6 %) (ex. **dépouiller l'arbre de Noël**, **description de poste**, **se faire frapper**, **filer mal** et **saison régulière**); enfin, nous y avons repéré 210 expressions figées (soit 23,44 %) (ex. **le diable est aux vaches**, **être à la fine pointe de**, **mouiller à siaux** et **sacrer son camp**).

Il apparaît donc clairement que la typologie des canadianismes anglais et des canadianismes français est très différente : deux catégories — mots simples et mots composés — accaparent les canadianismes anglais, avec les mots composés comptant pour plus de la moitié, tandis qu'en français, les quatre types d'unités sont représentés, avec les mots simples en dominance sur les autres.

Nous croyons que ces différences quant aux types d'unités lexicales entre l'anglais canadien et le français canadien s'expliquent également par l'évolution de ces deux langues au Canada. Le contact continu avec la Grande-Bretagne et les États-Unis qui a caractérisé le développement du Canada anglais fait que l'anglais canadien a senti surtout le besoin d'unités lexicales lui permettant de désigner des réalités concrètes caractéristiques du pays : par conséquent, il a eu recours à des mots simples et des mots composés. Le français canadien, d'autre part, isolé du français hexagonal par l'histoire, a acquis des particularités non seulement dans sa façon de désigner les réalités typiquement canadiennes, mais aussi dans la manière de s'exprimer à des niveaux plus abstraits, ce qui expliquerait la répartition des canadianismes français sur les quatre types d'unités.

6. TYPES DE CONCEPTS DÉSIGNÉS PAR LES CANADIANISMES ANGLAIS ET FRANÇAIS

Pour confirmer notre impression que les canadianismes anglais représentent surtout des réalités concrètes, tandis que les canadianismes français désignent aussi bien des réalités abstraites que concrètes, nous avons examiné les 50 premiers mots simples et composés

⁶ Nous tenons à préciser que, dans quelques cas, notre classification des unités lexicales étudiées pourrait porter à discussion. En effet, il arrive que la distinction entre deux types d'unités complexes ne soit pas évidente. Ainsi, nous avons parfois considéré qu'une unité lexicale était une collocation alors que d'autres auraient tout aussi bien pu juger qu'il s'agissait d'une expression figée par exemple. Mais nous ne croyons pas que ces quelques cas affectent sérieusement les résultats obtenus.

canadiens (ou avec un sens canadien) qui commence avec *D* dans le *GAGE* et le *RQ* (voir annexe)⁷.

Nous avons d'abord analysé ces mots selon leur degré d'abstraction. Nous les avons divisés en deux catégories : mots désignant des concepts concrets (objets, personnes, lieux, actions et qualificatifs concrets), et ceux désignant des concepts abstraits (idéas, émotions, états, activités intellectuelles, morales, ou émotives, qualificatifs abstraits, etc.). Selon cette catégorisation assez rudimentaire, 47 canadianismes anglais sur 50 désignent clairement des concepts concrets (ex. **deputy minister**, **dogsled**, **dew-worm**). Les trois qui restent semblent plus abstraits par rapport aux autres : **darb** («any thing or person thought to be especially large, good, etc.» - *GAGE*); **disallow** («of the Federal Government, nullify an act of a provincial legislature» - *GAGE*); et **disallowance** (1. «the power of the Federal Government to annul provincial legislation»; 2. «an act of the Federal Government exercising this power» - *GAGE*). Mais il ne s'agit pas d'un degré très élevé d'abstraction. D'autre part, sur les 50 canadianismes français, 9 sont carrément abstraits : **damné** («exprimant la désapprobation» - *RQ*); **débile** («Iron. extraordinaire, exceptionnel, fantastique» - *RQ*); **débouler** («(personnes) faire irruption» - *RQ*); **déculotter** («causer la ruine de qqn, lui enlever tout ce qu'il possède» - *RQ*); **se déculotter** («ne pas compter à la dépense, ne pas agir de façon mesquine» - *RQ*); **définitivement** («certainement, sans aucun doute» - *RQ*); **dégêner** («faire perdre sa gêne, sa timidité à qqn» - *RQ*); **se dégêner** (perdre sa gêne, sa timidité⁸); **se dégot(t)er** (se trouver, se dénicher⁹). Ainsi, les canadianismes français examinés comptent au moins trois fois plus de mots abstraits que les canadianismes anglais examinés.

Deuxièmement, nous avons réanalysé les mêmes vocables pour voir combien d'entre eux représentent des réalités typiquement canadiennes. Trente canadianismes anglais sur 50 désignent des réalités évidemment canadiennes : **dag**, **Dall('s) sheep**, **dan**, **Persian darnel**, **debt**, **debtor**, **Dene (n.)**, **Dene (adj.)**, **Dene Nation**, **deputy**, **deputy minister**, **deputy returning officer**, **detachment**, **development road**, **dew-worm**, **diamond willow**, **dickey (dickie, dicky)**, **Digby chicken**, **disallow**, **disallowance**, **Dogrib (n.)**, **Dogrib (adj.)**, **Dominion (adjl.)**, **Dominion Day**, **don**, **doré**, **dory**, **draegerman**, **drum dance**, **dugout**. À cette liste, nous aurions pu ajouter quelques autres vocables comme **dalles**, **dogsled**, **dog-sledge**, **dog-train**, et **Doukhobo(u)r**, qui représentent des réalités principalement canadiennes. Ainsi, la proportion des canadianismes anglais dénotant des réalités canadiennes est élevée. Cependant, parmi les 50 canadianismes français, il n'y en a que deux qui désignent une réalité purement canadienne — **D.E.C.** et **délégation** — même s'il y en a d'autres, comme les trois termes reliés au hockey, qui pourraient être considérés comme représentant des concepts principalement canadiens. Cette analyse montre donc clairement que les canadianismes en anglais servent principalement à présenter les réalités canadiennes, tandis que ceux en français désignent très souvent des concepts qui ne sont

⁷ Il est intéressant de noter qu'en nous limitant à 50 premiers mots qui commencent avec un *D*, nous épuisons les canadianismes anglais qui commence par cette lettre, mais il nous reste au moins trois fois plus de canadianismes en français.

⁸ C'est nous qui avons formulé cette définition, car le *RQ* n'en donne pas une pour le verbe pronominal.

⁹ C'est nous qui avons formulé cette définition, car le *RQ* n'en donne pas une pour le verbe pronominal.

pas propres au Canada. En d'autres termes, les canadianismes anglais donnent une idée de la vie et du paysage au Canada, tandis que les canadianismes français, tout en remplissant aussi ce rôle, présentent souvent des réalités que le Canada partage avec d'autres pays, mais d'un point de vue canadien.

CONCLUSION

Notre analyse des canadianismes anglais et français présente un portrait du Canada anglophone et francophone. Le portrait du Canada anglais qui en ressort est celui d'un pays avec certaines particularités géographiques, politiques, historiques qui ont besoin d'être nommées, mais d'un pays avec un nombre limité de ces particularités, car la culture canadienne n'est pas tellement différente de la culture américaine d'une part, et de la culture britannique de l'autre. Le fait qu'il y a peu de canadianismes anglais récents est peut-être une indication du fait qu'il y a relativement peu d'innovations modernes typiquement canadiennes. Les canadianismes français nomment aussi beaucoup de particularités canadiennes. Mais, parce que le Canada français a longtemps été isolé de la France, ces vocables présentent aussi d'autres réalités, abstraites aussi bien que concrètes. En outre, la présence des canadianismes modernes en français est une indication du fait que les Canadiens français veulent participer à la vie nord-américaine, mais dans leur propre langue. Ainsi, le français canadien donne l'impression d'un peuple plus dynamique que ne le fait l'anglais canadien.

ANNEXE

Canadianismes anglais dans le *GAGE*

1. dag, n. - a heavy, flat, double-edged triangular blade, used by the Indians as a weapon and tool
2. dalles, n.pl. - a narrow stretch of river between high rock walls, characterized by whirlpools, rapids, and treacherous currents
3. Dall('s) sheep, n. - a white North American sheep... of the western mountains found from northern British Columbia to the Arctic Ocean.
4. dan, n. - in the North : 1. a sealskin used as a container for oil; 2. a buoy made of inflated sealskin or sheepskin sewn airtight, used as a mark in deep-sea fishing
5. darb, n. - any thing or person thought to be especially large, good, etc. (*slang*)
6. Persian darnel - a common and troublesome Canadian weed introduced from Asia
7. dasher, n. - the fence surrounding the ice of a hockey rink; the boards
8. deadhead, n. - a log or fallen tree partly or entirely submerged in a lack, etc., usually with one end embedded in the bottom
9. debt, n. - 1. a credit given at a trading post to hunters and trappers in the form of supplies to be paid for out of

- the next season's catch; 2. the amount of credit given; 3. in the North, credit taken at any store
10. debtor, n. - a person owing or taking debt at a trading store
11. décharge, n. - a shallow stretch in a water course, where it is necessary to unload a canoe, etc. in order to make way by tracking or paddling
12. decker, n. (Baseball) a catcher's mitt
13. decoy, v. - deke
14. deer, n. - caribou
15. deke, v. - (Hockey) a fake shot or movement intended to draw a defending player out of position (*slang*)
16. Dene n.pl. - the Athapascan Indian peoples of the Northwest Territories
17. Dene, adj. - of, having to do with, or designating the Dene
18. Dene Nation - the official organization representing the Athapascan peoples of the Northwest Territories
19. deputy, n. - a representative to or in certain assemblies. In Quebec, the members of the National Assembly are often called deputies
20. deputy minister - in Canada, a senior civil servant who acts as assistant to a cabinet minister
21. deputy returning officer - in Canada, an official appointed by the returning officer of a constituency to look after the procedure of voting at a particular polling station
22. detachment, n. - the smallest unit in the organization of the Royal Canadian Mounted Police or other police force. Some rural detachments of the RCMP have only one or two officers.
23. development road - in the North, a road or one of a system of access roads intended to help the exploitation of natural resources
24. dew-worm, n. - any large earthworm that comes to the surface at night when there is dew on the grass. Dew-worms are often used as fishing bait.
25. diamond willow willow wood having a diamond-patterned grain, resulting from an abnormal growth of the stems that may occur in any species of willow. It is prized for making lamps, walking sticks, ornaments, etc.
26. dickey (dickie, dicky) n. - a pull-over garment for the upper body, having a hood made of duffel or skins; parka; atigi
27. Digby chicken - a small smoke-cured herring
28. dime, n. - a coin of Canada or the United States, equal to one tenth of a dollar; a ten-cent coin
29. disallow, v. of the Federal Government, nullify an act of a provincial legislature
30. disallowance, n. - 1. the power of the Federal Government to annual provincial legislation; 2. an act of the Federal Government exercising this power
31. dogan, n. - a Roman Catholic, especially one of Irish background

32. Dogrib, n. - 1. a member of an Amerindian people who live in the Northwest territories... 2. the Athapascan language of the Dogrib
33. Dogrib, adj. - of or having to do with the Dogrib or their language
34. dogsled, n. - a sled that is pulled by dogs
35. dog-sledge, n. - dogsled
36. dog-train, n. - a sled pulled by a team of dogs
37. dome fastener - a metal or plastic fastener consisting of two parts, one with a small, rounded projection in the centre that snaps into a sock in the centre of the other
38. Dominion, adjl. - 1. in Canada, under the control or authority of the federal government; 2. in Canada, relating to the country as a whole; national in scope
39. Dominion Day - a national holiday commemorating the establishment of the Dominion of Canada in 1867; Canada Day
40. don, n. - in some Canadian universities and colleges, an official in charge of a student residence
41. doré, n. - walleye
42. dory, n. - 1. John Dory, an edible sea fish; 2 walleye; doré
43. Doukhobo(u)r, n. - a member of a 200-year old Christian sect originally from Russia that traditionally believes that every person knows what is right and must be guided by this knowledge rather than by an outside authority. Several thousand Doukhobors left Russia in 1898 and settled in western Canada
44. draegerman, n. - especially in the Maritimes, a coal miner trained in underground rescue work and the use of special oxygen equipment effective in gas-filled mines
45. dram, n. - (*Historical*) (Logging) a section of a timber raft, made up of several cribs lashed together
46. driveway, n. - a road, especially one that is lined with trees and lawns
47. drum dance - 1. an Inuit dance consisting of expressive body movements and gestures, performed to the accompaniment of drums and with the dancer or dancers singing as they dance; 2. any of various Indian dances accompanied by drums
48. dude, n. - in the Western parts of Canada and the United States, a city-bred person, especially one who spends a holiday on a ranch
49. dugout, n. - especially on the Prairies, a large excavation used to hold water collected there from the spring thaw and from rainfall
50. duplex, n. - 1. a building consisting of two dwellings under one rook, either side by side or one above the other; 2. one of the dwellings of such a building

Canadianismes français dans le RQ

51. dactylo, n.f. - machine à écrire (*fam*)
 52. dalle, n.f. - gouttière (*fam*)
 53. dalot, n.m. - 1. bordure latérale creuse de l'allée de quilles; 2. partie profonde de la gouttière - tuyau de descente de la gouttière
 54. damné(e), interj. - exprimant la désapprobation (*fam*)
 55. danse carrée, n.f. - quadrille
 56. dard, n.f. - fléchette
 57. dardage, n.m. - (Hockey) action de darder; son résultat
 58. darder, v.tr. - (Hockey) donner un coup à un joueur avec la pointe de la lame du hockey (*fam*)
 59. se darder, v.pron. réfl. - (vers, sur) se précipiter, se jeter, s'élancer sur, vers
 60. carré aux dattes, n.m. - pâtisserie faite de dattes cuites disposées entre deux couches d'une préparation à base d'avoine ou de gruau
 61. débarbouillette, n.f. - petit carré de tissu éponge avec lequel on fait sa toilette, on se débarbouille
 62. débarquer, v.intr. - descendre de (*fam*)
 63. débarrer, v.tr. - ouvrir ce qui est fermé à l'aide d'un mécanisme (verrou, taquet, etc.), spécialt ouvrir avec une clé
 64. débattre, v.intr. - palpiter, battre très fort d'une manière déréglée
 65. débenture, n.f. - obligation non garantie
 66. débile, adj. - Iron. extraordinaire, exceptionnel, fantastique
 67. déboîter, v.tr. - détruire, réduire en morceaux (*fam*)
 68. débossier, v.tr. - débosser (*fam*)
 69. débossage, n.m. - débosselage (*fam*)
 70. débouche, n.m. - Abrév. de débouche-bouteilles
 71. débouche-bouteilles, n.m. invar. - ustensile servant à ouvrir les bouteilles capsulées
 72. débouler, v.intr. - (personnes) faire irruption
 73. débrocher, v.tr. - enlever les agrafes qui attachent des feuilles (*fam*)
 74. débroucheuse, n.f. - dégraffeuse (*fam*)
 75. D.E.C., n.m. invar. - Abrév. de *diplôme d'études collégiales*
 76. décharge, n.f. - cours d'eau dans lequel un lac déverse son trop-plein d'eau, spécialt l'endroit d'un lac d'où part ce cours d'eau
 77. déclin, n.m. - matériau de revêtement extérieur dont les éléments (planches, tôles, etc.) sont disposés horizontalement de manière à se chevaucher partiellement les uns les autres; ce revêtement
 78. décrassage, n.m. - [(Tech) opération qui consiste à débarrasser la grille d'un foyer des matières non combustibles]¹⁰ le décrassage d'un poêle
 79. décret, n.m. - ensemble de règles définissant les normes de travail d'un groupe de salariés
 80. décrochage (scolaire), n.m. - abandon des études avant la fin de la période d'obligation scolaire

¹⁰ Cette définition est tiré du PLUS, car le RQ ne donne qu'un exemple d'usage.

81. décrocher (de l'école) v.intr. -quitter l'école
 82. décrocheur, -euse, n. - élève qui abandonne ses études
 83. décrochir, v.tr. - redonner une forme droite à (*fam*)
 84. déculotter, v.tr. - causer la ruine de qqn, lui enlever tout ce qu'il possède (*fam*)
 85. se déculotter, v. prom. réfl. - ne pas compter à la dépense, ne pas agir de façon mesquine
 86. déductible, n.m. - part d'un dommage qu'un assuré doit assumer (*fam*) (*Anglic.*)
 87. défensif, -ive, adj. - (Hockey) Joueur défensif, avant qui se préoccupe davantage de la défense que de l'attaque; (Sports) Un match défensif, où la défensive domine
 88. défensive, n.f. - (Sports) la défense
 89. définitivement, adv. - certainement, sans aucun doute (*fam*) (*Anglic.*)
 90. défoliateur, -trice, adj. - qui cause la chute des feuilles d'un arbre
 91. dégêner, v.tr. - faire perdre sa gêne, sa timidité à qqn
 92. se dégêner, v. prom. réfl. - [perdre sa gêne, sa timidité]¹¹
 93. se dégot(t)er, v. pron. - [se trouver, se dénicher]¹²
 94. dégouttière, n.f. - 1. (surtout au plur.) liquide qui dégoutte, spécialt l'eau; 2. eau de pluie qui tombe d'un toit, d'une gouttière; glaçon, glace fondante qui dégoutte; 3. eau qui s'infiltré dans une construction et qui dégoutte
 95. dégraffeuse, n.f. - petit instrument de bureau servant à enlever les agrafes
 96. dégriffer, v.tr. - enlever les griffes à un animal, spécialt au chat
 97. déjeuner, n.m. - repas du matin
 98. déjouer, v.tr. - (Sports) manoeuvrer la rondelle, le ballon... de manière à empêcher un adversaire de s'en emparer
 99. délégation, n.f. - représentation permanente d'une province auprès d'un état étranger; fonction de délégué - ensemble du personnel assurant cette mission; résidence du délégué, locaux et services.
 100. deltaplanneur, -euse, n - adepte du deltaplane

¹¹ C'est nous qui avons formulé cette définition, car le RQ n'en donne pas une pour le verbe pronominal.

¹² C'est nous qui avons formulé cette définition, car le RQ n'en donne pas une pour le verbe pronominal.

RÉFÉRENCES

1. Ouvrages théoriques

- DUBOIS, J. et C. DUBOIS (1971) : *Introduction à la lexicographie : le dictionnaire*, coll. «Langue et langage», Paris, Larousse, 217 p.
- GREGG, R. J. (1989) : «La lexicographie de l'anglais canadien», *Revue québécoise de linguistique*, 18-1, pp. 151-187.
- MALLORY, J. R. (1984) : *The Structure of Canadian Government*, édition revue, Toronto, Gage Educational Publishing Company, 472 p.
- POIRIER, C. (1980) : «Le lexique québécois : son évolution et ses composantes», *Stanford French Review*, IV-1/2, pp. 43-89.
- REY, A. et S. DELASALLE (1979) : «Problèmes et conflits lexicographiques», *Langue française*, 43, pp. 4-26.
- RIALLAND-ADDACH, V. (1995) : «Culture partagée et dictionnaire monolingue de français langue étrangère», *Études de linguistique appliquée*, 97, pp. 91-103.
- SVENSÉN, B. (1993) : *Practical Lexicography : Principles and Methods of Dictionary-Making*, Oxford/New York, Oxford University Press, 285 p.

2. Ouvrages lexicographiques

- Collins English Dictionary* (1986) : Glasgow, Collins.
- Concise Dictionary of Canadianisms* (1973) : Toronto, Gage.
- Dictionnaire du français canadien* (1990) : Toronto, Stoddart.
- Dictionnaire du français Plus* (1988) : Montréal, CEC.
- Dictionnaire québécois d'aujourd'hui* (1992) : Saint-Laurent, Québec, Dicorobert.
- Funk and Wagnalls Canadian College Dictionary* (1989) : Toronto, Fitzhenry and Whiteside.
- Gage Canadian Dictionary* (1983) : Toronto, Gage.
- Lexis. Dictionnaire de la langue française* (1987) : Paris, Larousse.
- Penguin Canadian Dictionary* (1990) : Markham / Mississauga, Penguin / Copp Clark Pitman.
- Petit Larousse illustré* (1991) : Paris, Larousse.
- Petit Robert* (1991) : Paris, Robert.
- Random House Webster's College Dictionary* (1991) : New York, Random House.
- Richesses et particularités de la langue écrite au Canada* (1980) : Montréal, Université de Montréal.

PROBLÈMES DE L'ÉTABLISSEMENT DE LA TERMINOLOGIE JURIDIQUE DANS UNE LANGUE SANS OFFICIALITÉ HISTORIQUE. LE CAS DU BASQUE

Antton ELOSEGI ALDASORO

Université du Pays Basque, Donostia, Espagne

LA RELATION DE LA LANGUE BASQUE AVEC LE DROIT

La relation de la langue basque avec le droit a des aspects que l'on serait tenté de qualifier de paradoxaux —si cela n'étaient pas le cas de tant de langues, petites et grandes. Les pays parlant basque ont développé, tout au long de leur histoire, un bon nombre d'institutions juridiques propres, tant de droit public (les *Biltzar* du Labourd, les *Juntas Generales* de la Biscaye ou le Royaume de Navarre), que de droit privé. On trouve même, de nos jours, des institutions spécifiquement basques de droit privé actives dans le système juridique espagnol, et dont le nom basque n'a jamais été traduit (comme le testament dit de '*hilburuko*'). Pourtant, le droit basque n'a jamais été écrit en basque par des Basques mais plutôt en latin, romance navarrais, occitan provençal, occitan gascon, espagnol ou français.

À côté de ce paradoxe historique, nous nous trouvons actuellement face à un autre paradoxe : celui d'être en train de créer pour une vieille langue un monde conceptuel et terminologique, celui du droit, dont la caractéristique la plus manifeste est censée être sa traditionnalité, le fait d'être intimement liée à la mémoire historique de la langue qui a véhiculé ce droit. Mais nous savons bien que notre cas n'a rien d'extraordinaire comme le prouvent tant de pays qui ont été gouvernés dans des langues autres que la leur propre¹.

1. QUELQUES DONNÉES SOCIOLINGUISTIQUES ACTUELLES

La langue basque est parlée dans sept provinces historiques, 3 françaises et quatre espagnoles, par une partie variable de la population. Quelque 700 000 personnes en tout sur une population totale de quelque 2 860 000 habitants, dont 2 640 000 en Espagne et 220 000 en France. Il reste un nombre résiduel de bascophones monolingues (30 000), les autres ayant une bonne ou une très bonne compétence en espagnol ou en français. C'est-à-dire que les bascophones actuels sont une minorité bilingue de 23 % sur le territoire historique basque.

¹ Voir le livre de J.L. Mark Cooray (1985) qui décrit les problèmes auxquels le Sri Lanka indépendant a dû faire face pour arriver à faire fonctionner son droit dans les langues autochtones récemment officialisées.

Le Pays basque espagnol étant une région fortement industrialisée, la population bascophone est, dans sa grande majorité, urbaine et industrielle, mais les noyaux les plus compacts à dominance bascophone sont constitués de petites villes et de villages mixtes ruraux-industriels, hors des grandes villes comme Bilbao, Pampelune, Saint Sébastien. Du côté français, le monde rural est plus fortement représenté.

Un aperçu historique nous offre une image plus nuancée de la réalité; la situation ayant fortement changé depuis que dans la partie espagnole la langue basque a été reconnue co-officielle. D'après les données officielles, corroborées par une enquête sociolinguistique en 1996², on compte quelque 100 000 nouveaux bascophones depuis les 15 dernières années, dans une population à démographie stabilisée. Ce renouveau quantitatif de la langue basque est doublé d'une forte extension des usages sociaux dont cette langue avait été exclue historiquement. C'est justement un champ d'usage nouvellement ouvert, celui du droit, qui fait l'objet de cette communication.

Ces données montrent l'inversion du processus historique de perte continue tant du nombre des locuteurs que des relations sociales où la langue était employée. Processus accéléré avec l'extension de l'alphabétisation et l'importance croissante de la langue écrite qui condamnent sans merci les langues à usage uniquement oral.

Les facteurs de récupération linguistique sont bien établis. L'école vient en tout premier lieu : aujourd'hui tous les enfants d'une bonne partie du pays reçoivent leur éducation totalement ou partiellement en langue basque, ou bien apprennent le basque. Les progrès dus à l'enseignement primaire et secondaire s'ajoutent aux effets d'une chaîne de télévision émettant en basque toute la journée ainsi qu'aux moyens de communication en général, avec un quotidien et plusieurs hebdomadaires monolingues basques. Le troisième pilier de la récupération est l'emploi de la langue dans l'administration et à l'université; nous en parlerons plus tard.

Il reste tout de même de grands «trous noirs» où le basque n'a guère de présence, notamment dans la vie économique, les entreprises et les services privés et aussi, ce qui nous concerne plus directement, dans l'administration de la justice.

2. LA LANGUE BASQUE. DEUX MOTS SUR SON HISTOIRE ET SUR SON LEXIQUE

Du point de vue typologique le basque ou *euskara* est la seule langue non indoeuropéenne de l'ouest de l'Europe. Elle n'a pas de liens génétiques connus³ même si l'on a essayé de lui trouver des cousins chez nos voisins les Berbères ou chez les Kartveliens de Georgie ou bien, plus loin, chez les Aïnos du Japon.

Du point de vue historique, bien que de très sérieux indices lui donnent une ancienneté de quelques milliers d'années sur son territoire, son existence dans le Sud-Ouest

2 II^e Enquête sociolinguistique, 1996, Département de la culture du Gouvernement basque, Direction générale de politique linguistique du Gouvernement de la Navarre, et l'Institut culturel basque pour la partie française du territoire basque.

3 Trop de recherches et trop de pseudo-science ont été consacrées au «dévoilement des mystères» de l'origine du basque (voir J.L.Trask (1997) dans son dernier livre).

de la France et au Sud des Pyrénées depuis quelque dix-huit siècles est indéniable, puisque constatée épigraphiquement (Michelena 1984; Trask 1997). Ce territoire, jadis très vaste, est allé en se rétrécissant depuis la Rome Impériale, (Caro Baroja 1945) avec quelques rares épisodes d'expansion, sous la pression du latin et de ses langues filles. Mais il est toujours resté un noyau dur, qui comprend grosso modo le Pays basque français actuel (une moitié du département des Pyrénées atlantiques), plus, en Espagne, une bonne partie de la Navarre et de la région autonome qui porte le nom de Pais Vasco ou Euskadi.

Au cours de l'histoire, le Pays basque n'a jamais été totalement monolingue, puisque la langue d'écriture a toujours été autre — ce qui suppose une élite connaissant le latin et/ou d'autres langues romanes. De plus, entourée de langues plus «prestigieuses et cultivées», le basque a adopté un bon nombre d'éléments des langues voisines, modifiant progressivement, comme c'est le cas pour toute langue vivante, sa structure phonologique, sa syntaxe et surtout son lexique.

Le basque actuel garde dans son lexique de très importants champs sémantiques où la présence du basque primitif est manifeste. Mais, comme il fallait s'y attendre, les emprunts lexicaux au latin classique ou tardif, à l'occitan gascon et au castillan sont très nombreux et ce dans tous les champs «ajoutés» de la vie, depuis très longtemps. Ces derniers siècles, l'influence du français a été aussi considérable et, depuis quelques années, l'anglais commence aussi à laisser sa trace sur le lexique basque, à travers le filtre du français ou de l'espagnol.

Ce changement-enrichissement lexical prend une telle ampleur que la majorité des mots employés aujourd'hui dans le parler populaire, même rural, est étymologiquement non basque; leur adaptation phonologique et l'ancienneté de leur usage fait que les locuteurs n'ont souvent pas conscience de l'origine romane de ces mots de tous les jours⁴.

3. DU MANQUE D'OFFICIALITÉ À LA RECONNAISSANCE OFFICIELLE

3.1 L'histoire politique

La langue basque, paraît-il, n'a jamais été un véhicule écrit de la loi ni des relations juridiques privées, bien que différents territoires linguistiquement basques ont été politiquement souverains. La plus connue de ces entités politiques est le Royaume de Navarre, indépendant jusqu'au XVI^e siècle, à composant ethno-linguistique majoritaire basque, mais qui n'a jamais employé cette langue «officiellement».

L'histoire linguistique officielle des autres territoires, jouissant eux aussi d'un haut degré d'autonomie politique est semblable : leurs élites ont toujours préféré l'une ou l'autre langue néolatine.

4 Le lexique du basque actuel est un lexique «en voie de développement» assez avancée. La quantité et la variété des publications de différentes disciplines parues chaque année (1 305 en 1995, dont 467 classées 3-, Sciences sociales), et les douzaines de dictionnaires terminologiques publiés ainsi que des recherches comme la Collecte systématique du basque actuel (EEBS), corpus de 3 000 000 d'entrées lexicographiques en font la preuve.

L'histoire moderne du Pays basque est celle de la perte progressive de l'entité politique et de la fusion graduelle dans les grands États-nations, la France et l'Espagne, qui l'englobent. La perte de la personnalité politique et, donc, de l'aspiration à l'officialité apparaît comme définitive du côté français avec la Révolution⁵. Du côté espagnol, des guerres et des épisodes de violence, aux XIX^e et XX^e siècles, ont sûrement quelque chose à voir avec la résistance des Basques à la dilution de leur culture dans des ensembles uniformisateurs. Mise à part une courte expérience de 9 mois en 1936-1937, pendant la guerre suivant le soulèvement fasciste de Franco, où la langue basque fut déclarée officielle, jamais jusqu'à présent, c'est-à-dire jusqu'à la Constitution espagnole de 1978, la langue basque n'a eu de reconnaissance officielle.

On trouve, tout de même, des documents basques à valeur juridique depuis le XVI^e siècle : des témoignages transcrits littéralement, des traductions d'actes notariés, des registres communaux, des lettres entre des communes basques des deux côtés de la frontière, de très intéressants documents juridico-politiques des premières années révolutionnaires en France et, à partir du XIX^e siècle, au Pays basque d'Espagne, quelques ordonnances et des avis municipaux, ainsi que quelques contrats privés et d'autres documents commencent à être traduits en basque.

3.2 La situation actuelle dans l'administration et dans la vie juridique

En Pays basque de France le français reste seule langue officielle et valable; du côté espagnol, la Navarre, offre à l'*euskara* une demi-officialité dans une partie de son territoire. Les données qui suivent ne concernent que l'entité politique et administrative qui porte le nom de *Comunidad Autónoma del País Vasco* ou *Euskadi*.

Dans cette partie, la plus peuplée du pays (74 % de la population basque et 79 % des bascophones y habitent) et où les partis nationalistes sont au pouvoir régional depuis 1980, il s'est développé une politique d'application de l'officialité.

Ainsi, toutes les normes juridiques émanant des pouvoirs régionaux et locaux sont publiées en *euskara* en même temps qu'en espagnol. Chacun a le droit de s'adresser aux autorités dans la langue de son choix, et celles-ci sont (théoriquement) tenues de répondre dans la même langue. C'est surtout aux administrations locales des territoires à grande proportion de bascophones que la politique de basquisation de l'administration a été mise en vigueur⁶, faisant de l'*euskara* leur langue de travail. Mais pour la plupart des organes administratifs, le basque est une langue vers laquelle sont traduits des documents finaux dont l'instruction a été faite en espagnol.

5 La Révolution et la Grande Guerre ont marqué très négativement l'avenir de la langue basque en France, mais dans des différents épisodes de la Révolution l'*euskara* a eu plus de «reconnaissance» que sous n'importe quel autre régime. Voir Rica Esnaola (1975) et *1789 et les Basques* (1991).

6 Aujourd'hui, 28 communes entre 200 et 20 000 habitants se sont syndiquées dans le UEMA ou Mancomunauté des communes bascophones avec le but d'employer uniquement l'*euskara* comme langue d'administration. Seules les communications dirigées vers des non bascophones sont traduites en espagnol.

La politique de *capacitacion* linguistique (langue, traduction, rédaction) des fonctionnaires publics menée par l'IVAP (Institut basque de l'administration publique) et l'exigence de la connaissance de cette langue pour certains postes ont permis au bilinguisme de l'administration de faire des progrès.

La situation dans la vie juridique privée n'est pas si avancée. C'est surtout des entreprises de services publics, des caisses d'épargne et des compagnies d'assurances qui offrent leurs formulaires en format bilingue. Bien qu'officiellement valables, peu de contrats ou de testaments se font en *euskara*.

À l'université, les études de droit en basque se sont peu à peu développées depuis une dizaine d'années; aujourd'hui les deux facultés de Droit de la Communauté Autonome offrent la possibilité de faire une Licence en Droit en langue basque. Plus de 15 % des étudiants de la plus importante d'entre elles, l'Université du Pays basque, choisissent la «ligne basque» où la loi — presque toujours écrite en espagnol — est enseignée en basque. Les études juridiques de troisième cycle viennent juste de commencer : une thèse doctorale juridique a déjà été écrite en basque.

4. PROBLÈMES DE L'ÉTABLISSEMENT DU LANGAGE JURIDIQUE EN BASQUE

L'arrivée de l'officialité a supposé le passage d'une situation de vide presque total de langage juridique à une nouvelle situation où l'on produit chaque jour de nombreux textes juridiques, c'est-à-dire, des textes en langage juridique. Ce changement a demandé la résolution de différentes sortes de problèmes, concernant soit le langage du droit lui-même, soit la situation de la langue basque quant à son statut et quant à son corpus.

4.1 Problèmes concernant le langage du droit

De nombreux auteurs ont décrit (Cornu 1990) les traits caractéristiques du langage du droit et de sa terminologie. Je ne retiendrai que ceux qui supposent une difficulté ajoutée à l'établissement de nouveaux langages de droit.

- Le droit, comme système de régulation de la société, est tout d'abord un fait social et non pas une science, même s'il existe une science du droit. La terminologie du droit, préexistante à la science, n'a pas été créée ni structurée par des scientifiques.
- La terminologie du droit est dans sa grande majorité le résultat d'une spécialisation du lexique commun. La plupart des termes simples du droit sont des mots connus des simples usagers de la langue, mais le droit leur donne un sens spécialisé, quelquefois différent ou opposé à celui du langage commun.
- Le droit existe partout mais ses concepts ne sont pas universels. Beaucoup de termes juridiques designent des notions spécifiques des droits nationaux et sont intraduisibles *stricto sensu*, comme l'anglais *equity* ou les mots basques *hilburuko* et *elkarporoso*.

- Le droit est un fait social à grande continuité historique. Les changements politiques et même les révolutions sociales n'effacent jamais la continuité du vocabulaire traditionnel du droit.⁷
- Le droit est une réalité historique étroitement liée à la langue «nationale» de chaque pays. Il existe aussi de grands espaces de «culture juridique» supra-nationaux, mais la réalité «état-nationale» l'emporte toujours.
- Le droit, la loi écrite, a façonné, d'une certaine façon, les grandes langues porteuses du droit, les textes légaux ayant souvent été le modèle de prestige pour la langue écrite comme les *Leges XII Tabularum* ou les *Serments de Strasbourg*.
- La pratique juridique est souvent un exercice linguistique de commentaire de texte et d'interprétation, très difficile à faire dans une langue autre que celle du texte de la loi.

4.2 Problèmes de statut

· Une situation de minorisation linguistique : la population bascofphone est minoritaire, bilingue et habituée à employer le basque uniquement comme langue B, réservant à l'espagnol ou au français les usages à prestige social.

· De différents statuts légaux : le langage du droit n'est pas nécessaire du côté français; du côté espagnol, il a deux statuts différents : celui de la Navarre et celui de la Communauté autonome du Pays basque. La coordination d'ensemble reste, donc, impossible.

· Une officialité de deuxième classe. Quelques autorités en Pays basque, et en tout premier lieu les autorités judiciaires, ne reconnaissent pas dans la pratique le droit à la communication bilatérale en basque, notamment aux procès judiciaires. Une voie essentielle pour la normalisation du basque juridique reste ainsi fortement bloquée.

· Un usage symbolique de la langue : le basque est souvent employé non pas pour des besoins de communication mais comme symbole de la différence culturelle. Une partie importante des textes juridiques publiés en basque a rempli cette fonction; peu importe si l'on manquait de terminologie adéquate, ou si les textes traduits allaient être incompréhensibles pour les bascofphones.

· Un mauvais usage des recours : cet usage symbolique, quoique nécessaire, et le manque de planification adéquate, ont dévié des ressources humaines et financières qui auraient pu être mieux employées du point de vue de la normalisation linguistique.

4.3 Problèmes de corpus

· Une langue récemment unifiée : en 1978 quand la production de textes juridiques commence, la langue basque est loin de la standardisation. La *koiné* basque ou *euskara batua* crée par *Euskaltzaindia*, l'Académie de la langue basque, pour l'unification du basque écrit sur les dialectes, comptait à peine 10 ans.

⁷ Voir par exemple l'étude de Dubois (1962).

- Manque de dictionnaire du basque unifié, ce qui laisse aux rédacteurs juridiques — des traducteurs — le choix de beaucoup de mots du langage courant et, bien sûr, des termes juridiques. Par exemple, les traducteurs ont eu à utiliser intensivement un dictionnaire datant de 1905⁸, à grand intérêt philologique, mais qui avait écarté les mots jugés étrangers, bien qu'étant d'usage généralisé. Même si la situation a totalement changé ces dernières années⁹, les origines du basque juridique ont été marquées par le manque de base lexicographique suffisamment ample et acceptée.
- Manque de tradition juridique : les traducteurs ont eu à travailler sans modèle de référence propre, sur un langage fortement marqué par la tradition, tant pour le vocabulaire que pour la phraséologie, la syntaxe ou les différents registres linguistiques.

4.4 Problèmes concernant le corpus et le statut

- L'unité de la langue face au double système juridique : le basque unifié étant conçu pour servir de véhicule unitaire aux Basques de France ou d'Espagne, le langage créé pour le droit devait répondre en même temps aux besoins des deux systèmes juridiques. Problème sans solution dans beaucoup de cas, car les institutions juridiques ne sont pas identiques des deux côtés de la frontière. Par exemple, le basque (d'origine romane) *prokuradore* signifie *procureur* du côté français et *avoué* du côté espagnol.
- Un langage produit par des traducteurs : tandis que d'autres variétés ou technoclectes basques se sont développés, à partir surtout de l'école et des manuels scolaires, le langage du droit a été l'oeuvre des traducteurs du Parlement et des administrations publiques.
- Le basque est la langue du bout de la chaîne : le processus de production d'un texte se déroule entièrement en espagnol jusqu'à la phase finale où le traducteur intervient pour que le texte soit publié ou arrive chez son destinataire dans les deux langues officielles.

5. PRODUCTION DE TEXTES JURIDIQUES ET DE DICTIONNAIRES

5.1 Production de textes juridiques

Comme nous l'avons écrit, le langage juridique basque fut d'abord un langage créé par des traducteurs de l'administration publique. Depuis, la production de textes s'est diversifiée : la rédaction directement en langue basque se développe et ces textes appartiennent à des secteurs de plus en plus variés.

8 Le dictionnaire trilingue *Diccionario Vasco-Español-Francés* de R.M. Azkue est resté la référence principale jusqu'aux années 1980.

9 La lexicographie basque a fait des grands pas en avant. En plus de l'EEBS et des dictionnaires terminologiques cités, on dispose maintenant du volumineux dictionnaire général basque (OEH) (Michelena 1987) avec un corpus de 5 800 000 entrées lexicales qui reflète le basque des tous les textes traditionnels, partiellement publié, et l'on trouve sur le marché des dictionnaires de la langue actuelle comme le *Euskal Hiztegia* (Sarasola 1996) ainsi que des dictionnaires encyclopédiques en plusieurs volumes, en plus de dictionnaires bilingues (espagnol, français, anglais). *Euskaltzaindia*, l'Académie de la langue basque, prépare son dictionnaire du basque unifié qui supposera un pas très ferme dans la normalisation de l'*euskara*.

Ainsi, dès la fin des années 1980, l'*euskara* commence à être une langue d'enseignement du droit : des professeurs et des étudiants sont les nouveaux acteurs de la production de textes, ici aussi traduits dans la plupart des cas. Un certain nombre de manuels universitaires de droit les plus connus ont été traduits de l'espagnol, déjà quelques-uns ont été directement écrits en basque, et la production d'articles scientifiques sur des sujets juridiques commence à être normale dans des revues bilingues¹⁰. Il existe même une revue juridique rédigée exclusivement en langue basque¹¹.

5.2 Terminographie et normalisation du langage juridique

Au début, c'est les traducteurs eux-mêmes qui ont eu à jouer le rôle de terminologue et de terminographe, en assumant la responsabilité des choix terminologiques et en faisant des recopiations de termes. Vu la diversité des choix, des coordinations se sont créées assez tôt, mais bientôt on a pu constater que faute d'une autorité unique, les modèles linguistiques devenaient de plus en plus divergents.

Deux types d'institutions ont eu dès le début une incidence importante sur l'établissement du basque juridique : d'une part l'IVAP, déjà mentionné, et d'autre part, l'Institut basque de services universitaires (UZEI), institut privé à but non lucratif dédié à la confection de dictionnaires terminologiques; bien plus tard une troisième institution, l'université, s'y est ajoutée.

C'est UZEI qui en 1985 a publié le seul dictionnaire général de droit existant aujourd'hui. Ce dictionnaire de droit est l'oeuvre la plus marquante de la terminologie juridique basque, bien que contesté par un courant¹². Ce dictionnaire et la presque soixantaine d'autres dictionnaires spécialisés édités par UZEI représentent le fondement de la base de données terminologiques *Euskalterm* accessible «en ligne» et mise à la disposition de toutes les administrations publiques.

L'Institut basque d'administration publique à travers ses sections de traduction, terminologie et standardisation du langage administratif a surveillé la traduction des textes légaux et a surtout produit des manuels pour la rédaction et la traduction de textes administratifs ainsi que des recueils terminologiques partiels.

Quant à l'université, même si des universitaires ont toujours été consultés pour l'élaboration des terminologies, elle est arrivée depuis peu de temps à la production de recueils terminologiques en collaboration avec l'UZEI ou l'IVAP.

10 En tout cas, le droit est une discipline en retard sur beaucoup d'autres à l'université. La comparaison du nombre des thèses doctorales juridiques en basque (une seule) avec les plus de 70 thèses de tout genre lues en basque à la seule Université du Pays basque semble le démontrer.

11 La revue juridique *ELERIA* éditée par la Société des Études basques *Eusko Ikaskuntza* ne publie que des articles originellement écrits en basque sur des sujets de droit et de linguistique juridique.

12 Ce dictionnaire, *Zuzenbide Hiztegia* (UZEI : 1985), en deux volumes, dont l'un est monolingue basque et encyclopédique et l'autre un vocabulaire trilingue basque-espagnol-français, avec plus de 10 000 entrées à partir du basque, a été considéré représentatif du courant B que nous décrivons plus tard. Les critiques adressées à son égard lui reprochent surtout d'avoir voulu établir une terminologie avant même qu'il existe un dictionnaire général de langue commune moderne.

6. DIVERGENCES ET ESSAIS DE SOLUTIONS

Faute d'une tradition de référence et d'autorité terminologique capable d'imposer un modèle, les traducteurs qui ont créé le langage juridique ont essayé de fournir une réponse aux problèmes décrits, mais leurs directions ont été assez divergentes. Puisqu'il n'y a pas de véritable vie juridique en basque qui mette à l'épreuve la validité des textes produits et les choix terminologiques, les divergences loin de se diluer sont devenus des courants. Les différends existants à propos du modèle de modernisation de la langue en général, assez idéologisés d'ailleurs, se sont manifestés très explicitement à propos du langage juridique et administratif.

Ces diverses tendances ont produit non seulement des modèles textuels différents, mais aussi des normes et des terminologies assez divergentes. Il y a des cas où différentes traductions d'un même texte légal n'ont que 20 % de coïncidences lexicales, et peut-être encore moins pour les autres aspects textuels.

Même au risque d'en faire une caricature, nous allons essayer de signaler très schématiquement les traits distinctifs des deux principaux courants, que nous appellerons courant A et courant B (Berasategi 1989 : 29-45).

Courant A :

- Il veut être plus près de la langue orale traditionnelle, même aux dépens de la précision technique.
- Il donne plus d'autonomie au texte basque par rapport à l'espagnol : ses traductions s'éloignent davantage de l'équivalence formelle.
- Il est plus puriste tant du point de vue morphosyntaxique que lexical : il s'attache davantage aux structures considérées «plus pures» ou «moins latines», et priorise toujours les mots appartenant au fonds lexical basque par rapport aux emprunts.
- Ses textes sont souvent difficiles à comprendre, mais ils remplissent «mieux» la fonction symbolique décrite, puisqu'ils sont censés être plus «purement basques».
- Il est souvent en relation avec les organes supérieurs de l'administration autonome basque.

Courant B :

- Il cherche à répondre rapidement aux efforts de modernisation et d'adaptation de la langue à de nouveaux usages.
- Il cherche un plus grand rapprochement avec le parler des moyens urbains, sans trop se soucier de l'emploi des emprunts lexicaux.
- Il rejette l'emploi purement symbolique de la langue.

Il est préoccupé par la précision technique, essayant de trouver des équivalents lexicaux pour tous les mots espagnols, sans trop distinguer les termes techniques et les unités figées par la tradition.

Cette tendance est implantée surtout dans les administrations municipales et dans la terminographie spécialisée.

Il faut quand même dire que dès le début des années 1990 les vents poussent à la convergence. Même si des points de discussion restent encore sans accord, tant la coopération pratique que les débats qui se déroulent mènent à des solutions communes. Il est des thèmes où l'accord paraît bien établi, comme par exemple, la critique à l'emploi purement symbolique de la langue, ou la nécessité de s'adresser aux bascophones actuels dans un langage actuel, et non pas dans une langue idéale qu'il faudrait leur apprendre. Depuis quelques années, des commissions mixtes, représentant tous les courants et une grande variété de secteurs (académiciens, philologues, terminologues, représentants des diverses administrations, etc.), ont travaillé à l'élaboration de recueils terminologiques, ou de formulaires administratifs et judiciaires, arrivant à des solutions de consensus approuvées par tous.

Mais si l'accord entre les tendances est tout à fait nécessaire pour trancher des différends ayant à voir avec les idéologies à propos de la langue, il est clair qu'un approfondissement des problèmes plus spécifiquement linguistiques aiderait beaucoup à trouver des solutions.

Il faut tenir compte de la grande dépendance du langage de spécialité juridique à l'égard du langage commun. Plus que dans n'importe quel autre technocette, la plupart des termes juridiques sont à l'origine des mots de la langue commune. Les travaux lexicographiques en cours sur le basque unifié fixeront, sans doute, une base beaucoup plus sûre pour la terminographie juridique, mais ils devraient tenir compte du «côté terminologique» de beaucoup de mots.

Quant à elle, la terminographie juridique doit passer — elle est déjà en train de passer — à une nouvelle phase. Les dictionnaires juridiques et administratifs existants sont pour la plupart normatifs plutôt que descriptifs; ce sont des recueils de propositions plutôt que des descriptions de la réalité. C'est bien compréhensible puisque le corpus existant au moment de leur confection était plutôt maigre. Mais nous avons actuellement une masse assez considérable de textes juridiques qui n'a jamais été dépouillée.

Des travaux de dépouillement sur corpus nous permettraient de connaître le degré d'acceptation des diverses propositions terminologiques faites par les dictionnaires existants et de commencer à étudier le poids des emprunts, des calques et des autres formations néologiques dans l'établissement de la nouvelle terminologie juridique. Actuellement, des travaux terminographiques partiels ont déjà emprunté cette voie de dépouillement de corpus.

À part cet aspect quantitatif, nous avons aujourd'hui une partie de la terminologie juridico-administrative qui est effectivement employée comme langue de communication et de travail. Des études sociolinguistiques à ce sujet nous permettraient de mieux connaître

le degré d'usage, d'acceptation, et donc, de compréhensibilité des termes proposés par les différents dictionnaires.

7. CONCLUSION

Il est clair que seules les langues qui ont joui historiquement d'une reconnaissance de la part des autorités politiques ont pu développer sans entraves les différents langages de spécialité. Il est aussi évident que le technolècte du droit est la variété linguistique la plus dépendante de la politique : pourquoi écrirait-on des textes juridiques s'ils n'ont pas de valeur officielle ?

Ainsi, la langue basque a perdu la mémoire de son droit, et seulement quand elle a conquis son officialité le langage et la terminologie du droit ont commencé à être nécessaires. Mais l'arrivée de l'officialité ayant été vécu comme une victoire et une libération, les textes juridiques ont eu souvent à remplir une fonction symbolique plutôt que pratique. La traduction est devenue la source la plus importante de production de textes juridiques, au détriment de la création propre, et donc, de l'autonomie de la langue. Pendant un certain temps, des réponses improvisées à des besoins considérés comme pressants ont configuré des modèles de langage et des terminologies assez divergents.

Mais actuellement, à l'administration et à l'université surtout, le technolècte juridique est en train de passer l'épreuve de la pratique. La fonction symbolique prendra l'espace nécessaire mais réduit qui lui correspond. D'autre part, la lexicographie basque a fait d'énormes progrès ces dernières années, ce qui permettra de trancher beaucoup de discussions sur le lexique standard commun employé dans les textes juridiques. Il reste que les travaux terminologiques actuels, basés dans une meilleure connaissance du corpus existant, permettent de fixer une base plus ferme pour ce langage juridique que nous voulons authentiquement basque et pleinement communicatif et technique.

RÉFÉRENCES EN LANGUE NON BASQUE

- 1789 et les Basques* (1991) : Bordeaux, Presses Universitaires de Bordeaux, pp. 171-194.
- AGIRRE GANDARIAS, S. (1992) : «La oficialidad del euskera en documentos postmedievales». *ASJU*, XXVI,I, San Sebastian, Seminario de Filología Vasca Julio de Urquijo, pp. 259-279.
- AGIRRE GARAI, J. (1988) : «El lenguaje administrativo en euskera : situación y perspectivas», *Revista de llengua y dret*, 11, pp. 91-98.
- AZKARATE, M. (1988) : «Neologisms and Lexicography in the Basque Language», *Budalex Proceedings*, Budapest, Akadémiai Kiadó.
- BERASATEGI, J.I. (1989) : «La traducción jurídico-administrativa en el País Vasco», *Revista de Llengua i Dret*, 13, pp. 29-45.
- CARO BAROJA, J. (1945) : *Materiales para una historia de la lengua vasca en relación con la latina*, San Sebastian, Universidad de Salamanca.

COORAY, L.J. Mark (1985) : *Changing the Language of the Law. The Sri Lanka experience*, Québec, Les Presses de l'Université Laval.

CORNU, G. (1990) : *Linguistique juridique*, Paris, Montchrestien.

MICHELENA, L. (1984) : *Sobre la historia de la lengua vasca*, San Sebastian, Seminario de Filología Vasca Julio de Urquijo.

RICA ESNAOLA, M. (1975) : «Traduction en basque de termes politiques sous la révolution», *ASJU*, IX, pp. 3-172.

TRASK, L. R. (1997) : *The History of Basque*, London & New York, Routledge.

URRUTIA, A. (1987) : «Derecho, planificación lingüística y euskara», *HAEA/RVAP*, 17, pp. 119-123.

DICTIONNAIRES

AGIRRE, J. et A. DIAZ DE LEZANA, (1991) : *Vocabulaire des traités européens/ Europako Itunen Hiztegia*, Bilbao, IVAP.

AZKARATE, M., X. KINTANA et X. MENDIGUREN BEREZIARTU (dir) (1996) : *Elhuyar Hiztegia*, (dic. Basque-espagnol / Espagnol-Basque), Usurbil, Elhuyar.

MICHELENA, L. (1987) : *Diccionario General Vasco/ Orotariko Euskal Hiztegia*, Bilbao, Euskaltzaindia, Desclée, Mensajero.

SARASOLA, I. (1996) : *Euskal Hiztegia*, Donostia, Kutxa Fundazioa.

XARRITON, P. et X. KINTANA (1997) : *Dictionnaire/Hiztegia. Euskara-Frantsesa. Français-Basque*, Donostia, Elkar.

UZEI (1983) : *Administrazioa, Zirkulazioa Hiztegia*, [Dictionnaire d'administration et de circulation], Donostia, Elkar.

UZEI (1985) : *Zuzenbidea Hiztegia*, [Dictionnaire de droit], Donostia, Elkar.

UZEI (1994) : *Hirigintza Hiztegia*, [Dictionnaire d'urbanisme], Donostia, Elkar.

ORGANISATION STRUCTURALE ET SÉMANTIQUE DE LA MÉTAPHORE EN SONAY-ZARMA

Chaïbou NIANDOU

Université de Niamey, Niger

À propos de la notation

La notation que nous avons adoptée fait le plus souvent référence à l'Alphabet Phonétique International (API) et à la convention¹ de Bamako (du 28 février au 5 mars 1966). Le soNay-zarma est une des «langues» nigériennes concernées par les propositions de Bamako.

En effet, excepté les unités linguistiques ci-après² qui semblent présenter des réalisations moins originales, tout le reste doit être considéré comme conforme à la notation courante.

1. INTRODUCTION

La diversité des formes métaphoriques et leurs interprétations dans les interactions verbales en soNay-zarma est plus qu'une simple affaire d'analyse linguistique. En effet, la métaphore est une véritable ressource communicationnelle dans la vie quotidienne, en cela

¹ Cette convention a permis au gouvernement nigérien de prendre l'arrêté n° 017/MEN/ALPHA du 27 avril 1967 relatif à l'orthographe des langues nigériennes.

² Notations

- /N/ est réalisée nasale vélaire sonore (ex. haN «manger»)

- /ʌ/ est réalisée nasale palatale sonore (ex. ^â «mère»)

- /c/ est réalisée fricative palatale sourde (ex. céndí «tirer») et s'oppose à /j/, fricative palatale sonore (ex. jéndí «distance ou éloigné») et à /y/ qui est réalisée semi-voyelle palatale sonore (ex. yéndí «ramper, dans le cas d'une plante»)

- /u/ est réalisée voyelle postérieure premier degré d'aperture; elle peut être brève /u/ (ex. gúrí «puiser de l'eau») ou allongée (/u:/ ex. gú:rú «métal ou fer»)

Toutes les voyelles peuvent être réalisées nasales sauf /u/

- /a/ ex. há («demander ou force surnaturelle») /i/ : ex. hí «prêter»; - /o/ ex. hó «tiens!»; /e/ ex. hé «pleurer». Pour marquer une voyelle nasalisée nous utilisons le symbole appelé tilde /~/ . Le fait qu'il soit noté en dessous de la voyelle est une simple question pratique. En effet les tons ayant été placés au dessus de la voyelle, ajouter le tilde à la même place ce serait encombrant (cf. Hamani 1981 et 1983; Niandou 1993).

- /' / est réalisé ton haut (ex. bá «grande quantité») et s'oppose à / / réalisé ton bas (ex. bà «part»), à /ˀ /, réalisé ton descendant (ex. bâ «aimer») et à /_ /réalisé ton montant.

que les locuteurs de cette variété linguistique fondent fréquemment leurs façons de parler sur différents mécanismes de rapprochements entre certains signes linguistiques. Ces rapprochements, qui renvoient le plus souvent à un rapport d'analogie, donnent lieu à des formations de mots ou d'expressions, élargissant ainsi les champs communicationnels de la langue. On voit dès lors que la métaphore joue aussi un grand rôle dans ce qu'on a coutume d'appeler «création lexicale». Nous ne ferons pas ici une analyse approfondie de la question.

Notre démarche dans cette réflexion consiste à présenter quelques exemples de métaphores, dans une approche morpho-sémantique³ en tenant compte naturellement du fait que toutes les formes de métaphores ne pèsent pas le même poids.

Cette communication vise entre autres à dégager l'organisation interne (forme et contenu) de la métaphore en soNay-zarma. Nous tenons à préciser au passage que tous les exemples du corpus sont des formes linguistiques produites dans une région bien déterminée⁴. C'est le parler que nous utilisons couramment en famille et parfois avec les amis. Il se distingue des autres parlers du même continuum par certaines de ses variantes phonématiques et par une de ses marques prosodiques : l'accent⁵.

2. LA MÉTAPHORE : DES FORMES ET DES SENS

La métaphore peut se manifester sous trois formes principales en soNay-zarma :

1. métaphore à travers des composés à deux ou trois termes;
2. métaphore à travers un énoncé ordinaire (Fonction-Nature-Fonction);
3. métaphore à travers un énoncé proverbial.

2.1 Métaphore composée à deux termes

Gándà-kòrfú (litt. : corde de terre) «Serpent»

[Il s'agit ici d'une métaphore animalière, née d'une superstition. En effet, le serpent est considéré comme un génie. Il suffit d'en parler pour qu'on le voit. Or il est tellement dangereux qu'il vaut mieux ne pas le voir. Ainsi pour en parler sans qu'il le sache il faut lui trouver un surnom, mais qui colle bien avec sa morphologie. D'où celui de corde de terre.]

³ Nous entendons par morpho-sémantique, l'étude de la formation des mots ou expressions et de leurs significations.

⁴ Il s'agit d'une variante soNay parlée dans les cantons de Namaro et Karma (arrondissement de Kolo).

En règle générale, il semblerait que cette variante se rapproche des parlers soNay-ka:do des arrondissements de Tera et Tillabéri en ce qui concerne son «accent». De même qu'elle se rapproche des parlers zarma (Dosso, Niamey, Ouallam et Kolo) par son système morphosyntaxique, mais qu'au plan purement phonématique, elle se rapproche tantôt du soNay cité plus haut tantôt du zarma.

C'est donc pour sa position géographique et sa situation linguistique caractéristique d'une forme de transition que nous lui avons attribué le nom de soNay-zarma.

⁵ Nous entendons par accent un indice d'identification régionale.

^â-wà:kôy (litt. : celle ou celui qui a le lait de la mère) «Sorcier mangeur d'âme»

[C'est une métaphore à valeur de surnom qui vient de ce que la sorcellerie s'hérite par le lait maternel. Un individu dont le père est sorcier ne le devient pas lui-même : il est considéré seulement comme un homme méchant. Mais si sa mère est sorcière, il le devient automatiquement si elle l'allaita ou si une autre sorcière l'allaita.]

Bà:bà-fùmbó (litt. : père-pourri)

[«Ce qu'un homme représente pour le(s) enfant(s) de son épouse dont il n'est pas lui-même le père». C'est, en d'autres termes, le mari de la mère ou encore le beau-père. Cette entrée de pourriture traduit un rejet, et dans les deux cas d'ailleurs. Le beau-fils rejette son beau-père et inversement.]

ízè-fùmbó (litt. : fils ou fille-pourri(e)) «fils de la femme ou le beau-fils»

[Ce qu'un enfant représente pour le mari de sa mère dont il n'est pas lui-même le fils ou la fille.]

Wòybòrò-malfà (litt. : femme-fusil)

[«C'est une métaphore qui peut avoir différentes connotations selon le contexte. On peut l'attribuer à une femme courageuse qui, comparée à une balle de fusil, n'a peur de rien. On peut également l'utiliser pour une femme dangereuse par analogie fonctionnelle directe avec le fusil. On peut enfin l'utiliser pour désigner une très belle femme».]

Hási-tò:sí (litt. : caca de chien) «un petit vaurien»

[«Selon le ton utilisé ou le lien de parenté entre les interlocuteurs, cette métaphore prend des connotations plus ou moins péjoratives.

Quand des gens liés par une parenté se l'attribuent par plaisanterie, elle est moins péjorative et moins frustrante quel que soit le ton utilisé, car elle ne fait que renforcer davantage le lien. Lorsque cette métaphore est énoncée par les parents (mère, père, oncles ou tantes...) en direction de leurs descendances, elle contient toujours une marque de menace. Le ton est de ce fait dur et sérieux et a une valeur d'avertissement pouvant aboutir à une correction corporelle si l'enfant ne fait pas attention. Dans tout autre cas d'énonciation, elle sera considérée comme une injure».]

Kó:rò-kwá rày (litt. hyène blanche)

[«C'est une métaphore animalière à plusieurs connotations. On peut l'attribuer à une vieille personne ayant les caractéristiques physiques et comportementales d'une jeune personne. Elle peut désigner également un «coureur de jupon». Se dit aussi de quelqu'un qui aime jouer des sales tours. Enfin, elle qualifie celui qui aime trop la facilité.

Dans ces quatre cas, deux images s'opposent pour donner les interprétations sémantiques de cette métaphore : l'image que représente la couleur blanche, symbole de la sagesse, opposée à celle représentée par un manque de sérieux. En effet, jouer des sales tours, courrir derrière des filles, chercher trop la facilité sont des attitudes qui ne riment pas bien avec les cheveux blancs. Or, généralement, les personnes qui tardent à vieillir ont tendance à se comporter comme de jeunes gens. Le choix de l'animal, hyène, repose sur le fait qu'elle est considérée partout comme le personnage négatif de nombreux contes. C'est le symbole de la méchanceté et de la rapacité, d'où son manque de sagesse. La métaphore, dans ce cas là, exclut la stupidité de l'animal et rend le contenu moins insultant».]

àlbòrò-déjì (litt. : homme-feu) «un homme courageux et vaillant»

Géjì-háw (litt. : attacher la brousse)

[«C'est le nom donné aux rituels, sacrifices et incantations qui visent à calmer les forces de la brousse, à empêcher les génies de nuire, avant une opération de chasse ou une saison agricole par exemple...».] (cf. Olivier de Sardan 1982)

Fû-génjì (litt. : pilier ou support de maison)

[«Se dit de quelqu'un sur qui repose l'espoir d'une famille. Il peut s'agir également d'un génie protecteur d'une famille».]

Yòw-tàNgàrà (litt. : natte d'étranger)

[«respect envers quelqu'un pour son statut de nouveau venu ou d'étranger».]

Ízè-fùtú (litt. : fils méchant, fâché)

[«Un enfant indigne, mauvais garçon, insolent, désobéissant...; un jeune qui ne suit pas les recommandations des anciens, qui n'en fait qu'à sa tête».] Cf. J.-P. Olivier de Sardan (1982 : 141).

2.2 Métaphore composée à trois termes

Tà:mú-kún-tóndi:zè (litt. : cailloux de l'intérieur d'une chaussure) Voir ízè-fùmbó «enfant-pourri»

[«L'image du cailloux introduit dans la chaussure traduit une certaine gêne. L'enfant de la femme ou le beau-fils n'est pas toujours considéré comme un membre de la famille, mais le mari s'efforce de l'accepter comme tel. Ce jeu auquel doit forcément jouer le mari, s'il tient à rester avec sa femme, est en réalité très gênant. Par ailleurs, pour ne pas le qualifier «d'enfant-pourri», expression qu'il comprendra facilement, on préfère le surnommer Ta:mu-kuna-tondi:ze métaphore moins accessible et donc moins humiliante pour les enfants».]

Wòy-bà-^áyè (litt. : femme-plus que-parent)

[«un homme qui attache plus d'importance à sa femme ou à sa petite amie qu'à ses parents».]

àr-bà-^áyè (litt. : homme-plus que-parent)

[«une femme ou une fille qui donne plus d'importance à son mari ou à son petit copain qu'à ses parents».]

HáNgá-sí-dí (litt. : on n'en attrape pas l'oreille) «arrière-petit-enfant, arrière-grand-parent»

Jíná-hàré-yòw (litt. : étranger d'anti-chambre) «beau frère direct»

[C'est une métaphore née de la relation entre un homme et le frère de sa femme. Dans les sociétés que nous connaissons bien, faire l'amour avec sa sœur directe (c'est-à-dire la sœur de même père et de même mère), est pour une question de moralité l'un des incestes les plus condamnables. C'est une situation absolument inimaginable. De ce fait, le frère de la femme peut passer autant de temps dans l'anti-chambre avec sa sœur. Cela n'inquiète outre mesure le mari de sa sœur. Mais des personnes comme les demi-frères ou

les cousins peuvent relativement constituer un danger, par conséquent ils ne doivent pas rentrer dans l'anti-chambre».]

Hàsân-`ndá-hìni (l'étymologie de cette métaphore reste toujours problématique. Cependant quelques hypothèses⁶ sont émises à cet effet) «Force d'alliance imprescriptible»

[C'est une métaphore issue de la parenté et la force en question réside dans le rapport entre enfant d'oncle maternel et enfant de tante paternelle. Son origine est le cousinage à plaisanterie. L'invoquer à quelqu'un c'est lui jeter un sort (mort ou folie) dont il ne se remettra jamais».]

2.3 Métaphore comme énoncé ordinaire

à nà ázáw zí (litt. : Il a donné un coup de patte à l'ouest) «Il est mort»

[Le rapport analogique vient de la position d'une dépouille mortelle dans la tombe. En effet, dans la tradition musulmane les jambes du mort sont placées vers l'ouest.

Cette métaphore est normalement utilisée pour qualifier la mort d'un animal. Le terme de coup de patte a une connotation très péjorative et de ce fait ne doit pas être utilisé pour les humains.]

à háNá gá sándí (litt. : son oreille est dure) «il est têtu»

à bòNo gá sándí (litt. : sa tête est dure) «il est têtu»

à mòyey gá bá:nù (litt. : ses yeux sont mous)

[«Se dit de quelqu'un qui éprouve de la pitié et qui pleure facilement devant le malheur des autres. Elle peut s'attribuer également pour quelqu'un qui pleure facilement quand on le gronde ou on lui fait du mal».]

à kòy Ngá bà:bà (litt. : il a suivi son père) «Tel père, tel fils»

à má:yó bú (litt. : son nom est mort) «Il n'a plus de valeur, ou il est en déclin»

à déjò bú (litt. : son feu s'est éteint) «Il n'a plus de valeur, ou il est en déclin»

à n'à hárí haN (litt. : il a bu de l'eau) «il est âgé»

[L'eau désigne ici, la pluie. Dans les sociétés rurales, l'âge est souvent compté en fonction du nombre de saisons des pluies qu'a vécu chaque individu. Boire de l'eau signifie donc vivre plusieurs saisons de pluies.]

⁶. Une hypothèse de Jean Rouch, citée par J.-P. Olivier de Sardan, renvoie Hasan-nda-hini à une devise «zammu» des génies «to:ru» qui dit «je me confie à Dieu, je me confie à Hasan, je me confie à Hini». Il parle ainsi de Hasan et Hini, mystérieux ancêtres des peuples soṅay et zarma dont la tradition a perdu le souvenir. Jean-Pierre Olivier de Sardan tente lui-même de rapprocher Hasan de hášêy «l'oncle maternel», et Hini de híni «pouvoir», hypothèse qui nous paraît plus plausible.

2.4 Métaphore comme énoncé proverbial

Kóro cè dà jíNgàráy (litt. : pattes d'hyène et mosquée)

[«Se dit de quelque chose d'absolument impossible, irréalisable».]

Cè dàN kà mòy kà:jí (litt. : gratter les yeux avec le pied) «Se dit pour tout geste crapuleux, insensé»

Zùráy dà túndè kà:jìyáN (litt. : courir et se gratter les fesses) «Faire plusieurs choses incompatibles à la fois»

Hàyni-wi:kô dà tù:siji (litt. : coupeur de mil et morceau de tige)

[se dit de quelque chose qui est très facile à obtenir ou à réaliser.

Le mil (une céréale) représente l'alimentation de base de plus de la moitié de la population nigérienne, il est cultivé un peu partout dans le pays. Il pousse en tige et finit par un épis. La tige est formée d'une succession de nœuds. Alors, c'est l'intervalle entre deux nœuds que l'on fend en deux pour avoir ce qu'on appelle : *tu:siji* (équivalent du papier hygiénique) et qu'on utilise après s'être soulagé. Dans un champs de mil, le nombre de *tu:siji* ne se compte donc pas. Comme quoi, c'est ce qu'il y a de plus facile à trouver pour un cultivateur en train de couper son mil.]

Kàbè `ndà mê (litt. : main et bouche)

[«Elle est utilisée pour qualifier une distance insignifiante, deux pôles très proches l'un de l'autre et dont la jonction peut se faire de façon régulière. Le rapport est ici une image rendue par l'action de manger. Dans la plupart des sociétés rurales soNay-zarma, on mange avec la main, c'est-à-dire que les doigts remplacent la cuillère ou la fourchette. La main et la bouche sont donc deux choses qui peuvent naturellement se joindre à tout moment».]

3. CONCLUSION

Le processus de création morphosyntaxique et/ou lexicosémantique par la métaphore repose en général sur des rapprochements symboliques (euphémismes, interdits, tournures anecdotiques, injures, etc.)

Dans certains cas de rapprochements, la figure (signifiant métaphorique) et la référence (signifié métaphorique) sont incompatibles ou inattendues. D'où le caractère parfois instable (cf. Diébou 1994) du signifié métaphorique. C'est notamment le cas des formes métaphoriques proverbiales. À ce niveau le degré de connotation de la métaphore est plus fort. Cela relève d'un niveau de communication moins collectif, moins intégré. Les usagers doivent faire appel à leur performance linguistique pour comprendre la finalité du message.

À l'opposé, les productions métaphoriques composées ou phrastiques (énoncés ordinaires) relèvent le plus souvent du discours commun, plus intégré et dont la valeur symbolique est plus dénotative. Il y a compatibilité entre la figure et la référence. Il suffit pour l' usager de faire appel à sa compétence linguistique pour comprendre la finalité du message.

Le caractère superficiel de ce travail, présentation d'un corpus, car c'est de cela qu'il s'agit, ne nous permet pas de rendre compte véritablement de l'organisation sémantique et culturelle de ces types de métaphores en soNay-zarma. Cependant, c'est un effort qui présente un intérêt scientifique certain en matière de communication. Il permet de mettre en évidence des mécanismes complexes d'interactions verbales. Et d'en situer les difficultés (formulations, contextualisation, interprétation) dans l'intercompréhension entre des interlocuteurs de diverses couches socio-culturelles en vue d'assurer une réussite de leur communication. Or, pour qu'une communication soit considérée par les interlocuteurs comme réussie, disait Nadia Revaz, «il faut que l'émetteur et le récepteur soient capables de comprendre le signifié, de comprendre le référent et de comprendre la finalité du message linguistique».

RÉFÉRENCES

- DIÉBOU, T. (1993) : *Les proverbes zarma : structure linguistique et implications sociales*, Mémoire de maîtrise présenté à l'Université de Niamey.
- HAMANI, A. (1981) : *La structure grammaticale du zarma*, Thèse d'État, Université de Paris VII.
- HAMANI, A. (1983) : *De l'oralité à l'écriture. Le zarma s'écrit aussi*, publications de l'INDRAP, n° 158, Niamey.
- NIANDOU, C. (1993) : *Problématique de la traduction du français en soNay-zarma*, Mémoire de maîtrise présenté à l'Université de Niamey.
- OLIVIER DE SARDAN, J.-P. (1982) : *Concepts et conceptions songay-zarma*, Paris, Nubia.
- REVAZ, N. (1993) : «Difficultés de communication linguistique et rôle des savoirs culturels», *Bil*, 13, pp. 7-16.

'NE M'APPELEZ PLUS X MAIS Y' OU L'INTERVENTION DU DÉNOMMÉ DANS LE PROCESSUS DE DÉNOMINATION

René TONDJI-SIMEN

GRESLET, Université de Montréal, Canada

INTRODUCTION

Le présent travail a pour objet d'étude les anthroponymes. Il nous permet de revoir le concept et les principes de la dénomination anthroponymique et d'analyser les anthroponymes sous différents aspects : genèse, origines, formes, types, dation, changements et motivations. Une attention toute particulière est portée à l'intervention du dénommé dans le processus de dénomination.

1. DÉNOMINATION ANTHROPONYMIQUE

1.1 Concept

Pour s'adresser ou se référer à une personne, il faudrait d'abord l'identifier, en lui assignant un attribut : un nom.

Le nom situe la personne dénommée dans le réseau des us et coutumes de la société, de l'époque et du milieu géographique et social auxquels elle appartient. Donner un nom, comme le dit Clémentine Faïk-Nzuji Madiya :

«c'est dire qui est la personne nommée [...] c'est situer par rapport au temps, aux événements, aux comportements; c'est rattacher la personne aux choses et aux êtres, visibles, et invisibles [...] c'est communiquer, converser, dialoguer avec les autres, envoyer des messages, informer sur ce qu'on a longtemps gardé sur le coeur [...] c'est éloigner le porteur du nom des forces maléfiques».
(Madiya 1991 : 11)

Bien que pratiquée au niveau de toute l'humanité, la dénomination n'obéit pas à une norme universelle. Elle fonctionne selon les régions, les époques et les sociétés.

1.2 Genèse et origines des anthroponymes

a) *Genèse*

L'acte de dénomination remonte aux origines de l'humanité. Comme l'indiquent les premiers noms, Adam et Eve, la dénomination ne comprenait qu'un seul élément. Dans le

système indo-européen, le phénomène de nom unique s'est maintenu jusqu'à l'époque franque (Lefebvre-Teillard 1990 : 11). Le nom unique pouvait être simple (Plato) ou composé, permettant de varier les noms en même temps que de rappeler la filiation. Le système romain passa du nom unique (Romulus) au nom composé d'un *praenomen* (prénom : Marcus) et d'un *nomen gentilice* (nom de la gens, porté par tous les membres d'un clan) devenu par la suite héréditaire (Claudius). Et comme les personnes qui portaient les mêmes *praenomen* et *nomen* étaient très nombreuses, on introduisit un autre nom héréditaire, le *cognomen*, pour distinguer les diverses branches d'un même clan (Cicero). Chaque individu pouvait en plus se choisir un *agnomen* ou surnom. Le nom romain pouvait donc se composer de quatre éléments (Publius Cornelius Scipio Africanus) (Britannica 1991 : 731, vol. 24). À la fin de l'empire, ce système de dénomination disparut car ceux qui portaient des noms qui n'avaient aucun rapport avec leur identité étaient de plus en plus nombreux (les esclaves affranchis et leurs descendants, les étrangers devenus citoyens). La diffusion et le triomphe du christianisme contribuèrent également au recul de ces noms païens, à la montée des noms chrétiens et à leur introduction dans des territoires lointains. Vers le XII^e siècle, le système des noms chrétiens s'était solidement implanté.

Les noms de famille firent leur apparition vers l'an 1037 quand les fiefs devinrent héréditaires et que la coutume d'ajouter le nom de la ferme au nom propre se répandit. Puisque la ferme était devenue héréditaire, le nom de la ferme devint nom de famille, nom donnant droit à la ferme familiale. Une branche de la famille prenait un nouveau nom de famille quand elle acquérait une autre ferme. Il fut ordonné en 1828 qu'une personne qui changeait de ferme ne devait plus abandonner son ancien nom, mais l'ajouter à celui de sa nouvelle ferme (Nell 1974 : 75). Ainsi s'installa, en Europe, le système de nom de famille.

En ce qui concerne l'Afrique, l'oeuvre de Willy Bal (1991 : 3) nous permet de comprendre les voies par lesquelles les anthroponymes romans ont pu s'insérer dans les usages de l'Afrique noire :

i) voie généalogique : noms romans donnés aux enfants issus d'une miscigénération et que le père d'origine romane installé en Afrique a reconnu (Sao Tomé, sud de l'Angola, Afrique du Sud, Namibie);

ii) voie pseudo-généalogique : au XIX^e siècle, après l'abolition de l'esclavage au Brésil, des esclaves africains affranchis qui avaient adopté les noms de leurs anciens patrons, sont venus s'établir sur l'ancienne Costa da Mina (Togo, Ghana, Bénin) et ont transmis à leurs descendants leurs patronymes : *Amorin, Da Silveira, De Almeida, De Campos, De Souza, Gomes, Medeiros, Olímpio, Paraíso, Sacramento, Santos* (Bal 1991 : 3);

iii) voie de l'acculturation : divers faits historiques (colonisation, évangélisation, etc.) ont introduit dans le répertoire anthroponymique africain des catégories linguistiques européennes (prénoms chrétiens, patronymes, titres, noms ethniques, professions, positions sociales). Exemples : *Senghor* venu du nom portugais *Senhor*; *Nzuji* du français *juge*; *Madiya* du nom portugais *Maria* (Bal 1991 : 4).

b) Origines

Les anthroponymes peuvent être d'origine mythologique, religieuse, païenne (noms des lieux géographiques où la famille réside ou a résidé).

Les noms peuvent dériver du nom du chef de la famille ou du géniteur (Fernandez, fils de Fernand; Johnson, fils de John) ou se composer d'une partie du prénom et du nom du père comme chez les Frisons, par exemple, où le fils de Thorn Jansen s'appelait Frerck Thorsen et le petit-fils Paul Frercksen (Nell 1974 : 7).

Ils peuvent aussi avoir leur origine dans le diminutif ou le surnom porté par le chef de famille (Livi Bacci 1974 : 84), provenir des numéros d'ordre comme les noms romains d'autrefois (*Primus, Secundus, Tertius*), des chiffres, des métaphores imagées (Dauzat 1977 : 10), des qualificatifs, des noms communs, des noms d'animaux, de plantes ou de déités.

1.3 Types et formes d'anthroponymes

Comme nous l'avons déjà mentionné, un anthroponyme peut se composer d'un, de deux, de trois, de quatre éléments et plus. Ces éléments portent des noms différents selon le paramètre et le système terminologique considérés : nom, prénom, nom de baptême, nom de confirmation, nom de naissance, nom de mariage, nom de famille, nom d'auteur, nom de guerre, alias, surnom, sobriquet, pseudonyme, nom ethnique, etc. Les types et les formes des anthroponymes varient.

a) Types d'anthroponymes

Clémentine Faïk-Nzuji Madiya (1991 : 12-23) classe les anthroponymes africains selon la typologie suivante :

i) D'après le moment d'attribution du nom :

- Nom de naissance : le nom de naissance est qualifié de «principal» (ou «nom» tout court) s'il sert à l'identification sociale et familiale, et de «deuxième» s'il sert à l'identification personnelle au sein de la famille :

«En d'autres termes, le "nom principal" correspond à ce qui, dans la terminologie européenne, est appelé "nom" et le "deuxième nom" à ce qui est appelé "prénom"». (Madiya 1991 : 12)

- Nom acquis : nom qu'une personne reçoit lorsqu'une circonstance spéciale l'y autorise : initiation, intronisation, consécration.

ii) D'après le mode et la fréquence d'utilisation :

- Nom courant.

- Nom privé : nom qui n'est pas nécessairement utilisé par tout le monde (1991 : 12).

- Nom latent : nom à caractère plus ou moins secret. Les noms acquis dans les confréries secrètes sont de cette catégorie (1991 : 12).

iii) D'après l'usage :

- Nom d'adresse : pour s'adresser à la personne directement (1991 : 12).

- Nom de référence : pour parler à quelqu'un d'autre d'une personne (1991 : 13).

iv) Noms de circonstances :

- Noms provenant d'événements particuliers ayant eu lieu avant la naissance, pendant la grossesse ou au moment de la naissance de ceux à qui on les donne (1991 : 15-16).

- Noms désignant le comportement des parents : L'Accueilli, L'Attendu (1991 : 16).

- Noms mettant en relief le comportement de l'entourage : On Me Hait, On A Déjà Médit de Moi (1991 : 16).

- Noms faisant état d'événements historiques et de phénomènes naturels s'étant produits pendant la grossesse ou la naissance; ou des lieux de l'accouchement : La Guerre, La Colline, La Variole, La Rivière (1991 : 16).

v) Noms de naissances «glorieuses» :

Ces noms donnent des renseignements sur le type de la naissance :

- Noms des jumeaux.

- Position de l'enfant par rapport aux naissances précédentes ou postérieures.

- Position du bébé à la sortie (1991 : 17).

vi) Noms de réincarnation :

- Noms liés à la croyance selon laquelle des ancêtres ou esprits incarnés défunts peuvent se réincarner dans de nouveaux corps humains (1991 : 18).

vii) Homonymes :

- Noms des membres de la famille, d'amis ou de chefs de lignages attribués aux nouveau-nés (1991 : 19).

viii) Éponymes, ethnonymes et noms de clan :

- Noms attribués aux enfants de même lignage, ethnie ou clan.

ix) Noms de pouvoir et de position sociale :

- Noms de titres honorifiques acquis après l'accession à un statut socio-politique particulier (1991 : 20).

x) Noms d'accomplissement :

- Noms acquis par le porteur lorsque, dans sa vie, intervient un changement important d'ordre physique, social ou spirituel autre que dans le cas du pouvoir politique. Ex. : une mère après la naissance de son premier enfant ou les parents de jumeaux (1991 : 21).

xi) Noms forts : noms de louanges, noms de tambour :

- Les noms forts sont des termes ou des syntagmes de louanges ou de tambour. Les noms forts sont une forme de devise, individuelle ou collective (1991 : 22).

xii) Noms de plaisanterie, surnoms ou sobriquets :

Ces noms traduisent les plaisanteries entre grands-parents et petits-enfants, les relations de camaraderie, les critiques, les moqueries (1991 : 23).

b) *Formes des noms anthroponymiques*

La forme des noms dépend des facteurs culturels et géographiques.

Dans beaucoup de pays occidentaux, la forme la plus fréquente est de type : Prénom + Nom de famille : François Mitterrand.

En Chine et en Hongrie la forme la plus courante est plutôt du type : Nom de famille + Postnom : Mao Tse-Tung, Nagy István.

Au Canada et aux États-Unis, entre le prénom et le nom de famille, on insert habituellement un autre élément, appelé deuxième prénom : Prénom + 2^e Prénom + Nom de famille : Pierre Elliott Trudeau, William Jefferson Clinton, John Fitzgerald Kennedy.

Habituellement, juste la première lettre du deuxième prénom est utilisée. Aux États-Unis, surtout, la forme hypocoristique des noms est souvent utilisée : William : Bill; Sammy, Samuel : Sam; Robert : Bob.

En Russie, habituellement, un nom patronymique (dérivé du prénom du père) est inséré entre le prénom et le nom de famille : Prénom + Patronyme (dérivé du prénom du père) + Nom de famille. Si, par exemple, un père se nomme Ivan Krylov :

- le fils peut avoir pour nom : Igor + Ivanovich + Krylov
- la fille : Varvara + Ivanovna + Krylova

En Espagne, habituellement, le nom de famille est formé du nom de famille du père et de celui de la mère : Prénom + Nom de famille du père + Nom de famille de la mère.

En anglais, beaucoup de noms de famille sont formés d'un patronyme ou d'une de ses variantes et d'un suffixe ou préfixe (*Son, Mac, Mc, Fitz, O', P-*) signifiant fils :

Suffixation : Patronyme + *son* : Johnson

Préfixation : *Fitz* + Patronyme : Fitzgerald

Mac ou *Mc* + Patronyme : MacBeth, MacDonald

O' + Patronyme : O'Brien (Fils de Brien)

P + Patronyme : Powell (Fils de Howell)

En grec : Suffixation : Patronyme + *poulos* (fils de) : Assimakopoulos

En bamiléké et en bassa, langues du Cameroun, le préfixe indiquant la filiation s'adjoit aux noms de fille : Ngo + Patronyme :

NgoNkuibo (Bamiléké-Medumba : fille de Nkuibo).

NgoBayiha (Bassa : fille de Bayiha).

En français :

a) Formation d'anthroponymes à partir d'un autre :

Jacques : Jaquet, Jacquot, Jacotot, Jacquart, Jacquier, Jacqueline (Britannica 1991 : 732, vol. 24).

b) Préfixation d'une préposition ou d'un article à des noms

- de structure ou de construction : Dupont, Dutoit, Dupuis, Dutemple, Demaison, Lachapelle, Lahaie, Laporte;

- de lieu géographique : Davignon, Decaen, Derennes;

- d'habitant : Lebreton, Lenormand, Lebasque;

- de fruit, de fleur, de plante ou d'arbre : Laframboise, Larose, Latulipe, Lavigne, Deschênes;

- de couleur : Lebrun, Leroux;

- de particularité géomorphologique : Lamontagne, Lavallée;

- de particularité physique ou morale : Lesage, Leborgne;

- de titre : Leduc, Lecomte, Leroy, Lenoble.

c) Transformation de titres professionnels en anthroponymes :

Boulangier, Masson, Carpentier, Boucher, Couturier.

d) Transformation de noms d'animaux en anthroponymes :

Leloup, Leboeuf, Chevrette, Cauchon.

En Afrique, le modèle européen s'est implanté presque partout. Dans la plupart des pays, les anthroponymes se composent d'un ou de deux noms africains et d'un ou de plusieurs prénoms ou postnoms occidentaux, chrétiens ou musulmans selon qu'on est en territoire à forte présence chrétienne ou musulmane. Les autres noms africains étant réservés aux communications intrafamiliales ou intra-ethniques. Toujours sous l'influence culturelle occidentale, des noms de famille s'implantent et comme ils ne peuvent pas faire disparaître les homonymes, ils s'adjoignent à eux. Ainsi, les anthroponymes à trois éléments sont de plus en plus courants :

Prénom + Homonyme + Nom du père : René Tondji Simen.

Quand l'homonyme est le nom du père, l'anthroponyme se réduit à deux éléments :

Prénom + Homonyme : Rosine Tchakountio.

Dans certains cas d'homonymie avec le père, on préfère répéter l'homonyme :

Prénom + Homonyme + Homonyme : Samuel Nganbe Nganbe.

La société arabe n'a pas échappé au modèle européen constitué d'un prénom et d'un nom :

- de filiation : Ali Ibn Abbas (Ali fils de Abbas);
- de lieu : Ibrahim Al-Baghdadi (Ibrahim de Baghdad);
- de profession : Mansur Al-Ghazali (Mansur le Fileur).

En Amérique espagnole, deux modes sont en usage :

- a) Prénom + nom du père + nom de la mère.
- b) Prénom + nom du père (Sanchez-Albornoz 1974 : 15).

Au Québec, l'enfant qui naît peut porter un ou plusieurs prénoms + le nom du père ou le nom de la mère ou le nom du père et celui de la mère dans l'ordre choisi par les parents et reliés par un trait d'union. Et quant au nom de mariage, la conjointe conserve son nom de famille.

2. DATION DE NOM

La dation de nom peut se faire par transmission, attribution, auto-attribution ou imposition. Elle est tributaire de la culture, du donneur et des motivations de ce dernier.

2.1 Culture

La dation des noms se fait selon la culture. Dans la plupart des cultures occidentales, le nom (nom de famille) se transmet directement du père aux enfants dans les cas de patronymie ou des parents aux enfants dans les cas des noms composés. Les prénoms sont attribués par les parents ou le prêtre ou le directeur de l'état civil.

Dans des cultures africaines, les noms de naissance (ou homonymes) sont attribués par les parents ou grands-parents en souvenir des gens qu'ils aiment. Les prénoms ou postnoms (prénoms de type occidental, noms de circonstances, noms courants, noms de réincarnation), quant à eux, peuvent être attribués par les parents, les prêtres, les autres chefs religieux, les amis de la famille ou l'officier de l'état civil. Les éponymes, les ethnonymes et les noms de clans sont transmis comme les noms de famille en Occident. Les noms latents, les noms de pouvoir et de position sociale, les noms forts sont attribués par les chefs et les notables. Les noms d'accomplissement et les noms de naissance glorieuse s'imposent automatiquement aux porteurs dès que les conditions sont réunies. Dans la culture bamiléké-medumba, par exemple, les jumeaux doivent porter des noms se terminant par *mi*, celui qui précède les jumeaux doit s'appeler *Mbatkam*, la mère des jumeaux *Magni* et leur père *Tagni*, le dernier-né de la famille *Kebatjap*, la femme qui vient d'accoucher *Mabwe*, la première femme d'un polygame *Mala'*, la dernière *Njuiguen*, la femme qu'un nouveau chef prend pendant la période d'internement au cours de laquelle les neuf notables supérieurs le préparent à l'exercice du pouvoir s'appelle *Njuikam*, et la première qu'il avait avant son accession au pouvoir *Makwa*, la reine-mère *Mamven*, l'aîné

du chef *Tamven*, le premier fils que le chef obtient après son accession au trône *Pokam*, le frère cadet immédiat du chef *Toukam*, l'adjoint au chef *Nkuibou*.

2.2 Donneur

De manière générale, le donneur de nom doit être une personne dont le pouvoir et l'autorité en cette matière sont reconnus dans sa communauté. Dans la culture occidentale, la question ne se pose que pour le prénom. Dans les cultures sans tradition de nom de famille, l'importance du donneur de nom de naissance est considérable. Étant donné que le nom ne s'attribue pas au hasard, donner le nom est un acte presque religieux réservé, très souvent, au chef de la famille. Un autre membre de la famille peut le faire si cette responsabilité lui est reconnue par la famille ou la tradition. Quant au nom de circonstances, la tradition est moins exigeante, mais très souvent, c'est la grand-mère ou la femme la plus âgée ou la plus aimée de la famille qui le choisit. Le prénom occidental est choisi par une des personnes scolarisées de la famille ou par le pasteur, le prêtre ou l'officier de l'état-civil. Chez les Dogon, par exemple, qui reçoivent trois noms, il y a trois donneurs : le patriarche donne le nom courant (nom officiel), la famille maternelle donne un nom privé qu'elle est, par ailleurs, la seule à utiliser, le prêtre donne un nom secret (Retel-Laurentin 1972 : 17).

Les titres honorifiques ou de notabilité sont donnés par les différents détenteurs de l'autorité traditionnelle : rois, chefs supérieurs et notables.

Les donneurs peuvent également être : a) un groupe d'amis ou des pairs pour les surnoms ou les sobriquets; b) des autorités militaires, pénitentiaires ou des services secrets pour les noms codés ou noms-numéros; c) des maîtres pour les noms des esclaves; d) des colons, des racistes ou des xénophobes pour des appellations péjoratives, ethniques, stéréotypées; e) des artistes qui se donnent des noms ou des pseudonymes.

2.3 Motivations

Les motivations des donneurs de noms sont de différents ordres allant des plus nobles aux plus abjectes.

Pour le choix des prénoms, les parents et les amis se laissent souvent guider par la mode, la beauté ou la rareté du nom, et le fait que certaines personnalités célèbres ou des gens symbolisant la réussite ou les qualités qu'ils aiment aient porté ces prénoms. Les prêtres catholiques donnent à l'enfant le prénom du saint du jour de naissance ou de baptême (Nell 1974 : 12). Les parents africains tiennent compte des circonstances ayant entouré la venue au monde du nouveau-né et de la morale sociale ou éducative qu'ils veulent inculquer à l'enfant et à leur entourage.

Les colons veulent propager leur culture et leur civilisation, les maîtres d'esclave marquer leur domination et assujettissement, les racistes et les xénophobes, leur haine et mépris.

Les noms d'artiste, les pseudonymes et les noms codés sont destinés à protéger l'identité de leurs porteurs. Les surnoms, quant à eux, sont utilisés plutôt pour mieux faire connaître la personne ou une de ses particularités et le plus souvent dans le but

d'augmenter sa popularité. Les surnoms des sportifs (lutteurs, boxeurs, basketteurs, footballeurs, toréadors...) ne remplissent que parfaitement cette fonction.

3 CHANGEMENTS DE NOM

3.1 Changement de nom par attribution

Dans le passé, quand on acquérait un domaine, on devait prendre le nom du domaine (Dauzat 1977 : 342).

Les maîtres imposaient à leurs esclaves des noms qui n'avaient aucun rapport avec leur culture, et le jour où ils les affranchissaient, ils leur attribuaient leurs propres noms et prénoms. Quand les Indiens voulaient devenir citoyens américains, ils devaient renoncer à leurs noms indiens et prendre ceux des Blancs (Adler 1978 : 138).

Un musulman, un païen, un Noir, un Asiatique ou toute autre personne qui se convertissait au christianisme recevait à son baptême un nom chrétien. Cette pratique est de rigueur encore aujourd'hui dans beaucoup de religions (Albert Bernard Bongo/Omar Bongo; Cassius M. Clay/Muhammad Ali).

Dans l'Église catholique, le pape change de nom quand il est élu. Dans la société Maaaga, chez les Mossi, un chef élu change également de nom. Et tous ceux qui portaient le même nom que lui avant son élection changent eux aussi de nom et s'appellent «Naab-Yuure» (Ilboudo 1977 : 107).

Chez les Bamiléké, l'enfant enlevé par le chef pour servir à la chefferie perd définitivement son nom. On ne l'appelle plus que par un titre correspondant à sa fonction. Quand il quittera la chefferie pour devenir notable, le chef lui attribuera un nouveau nom (Nissim 1977 : 58).

En 1972, au Zaïre, par exemple, une décision politique obligea les Zaïrois à remplacer leurs prénoms chrétiens par des noms «authentiques» (Madiya 1991 : 21). Joseph Désiré Mobutu, président de la République à l'époque, devint Mobutu Sese Seko.

Chez les Blackfeet, une tribu indienne de l'Amérique du Nord, traditionnellement, on doit changer de nom à chaque saison et quand quelqu'un a fait un acte de bravoure, il a droit à un nouveau nom (Adler 1978 : 122-132).

Chez certains peuples, on prend un nom à chaque étape de la vie : enfance, adolescence, quand on devient adulte ou se marie et à la vieillesse. Les Iroquois en prennent un nouveau quand ils deviennent adultes et par après à chaque fonction sociale qu'ils occupent (Alford 1988 : 6).

3.2 Changement de nom par autodénomination

Différentes raisons peuvent amener une personne ou une communauté à changer de nom : nom ridicule, malséant, vulgaire, déshonorant ou péjoratif, rectification d'un acte de l'état civil, naturalisation, mariage, divorce, décès du conjoint, adoption d'une nouvelle religion.

En raison du mariage, la conjointe peut prendre le nom de son mari ou l'adjoindre au sien par un trait d'union comme en France et au Canada; ou le mari peut prendre le nom de sa femme en respect du principe d'égalité comme en Tchéquie.

En cas de décès du mari ou de divorce, la conjointe peut reprendre son ancien nom ou s'en attribuer un nouveau.

Si quelqu'un adopte une nouvelle religion, il peut remplacer son ancien nom par un nom reconnu dans sa nouvelle religion.

Le changement de nom par autodénomination est une pratique assez courante chez les artistes. Plusieurs raisons peuvent le justifier : modestie, protection contre des critiques ou toute autre forme d'attaque, libre expression, mystification, intérêt à ne pas faire savoir de quel sexe on est (surtout quand les femmes-artistes n'étaient pas bien vues), changement de style, pratique journalistique (anonymat ou nom différent selon l'éditorial ou l'article) (Nuessel 1992 : 18). Des écrivains adoptent à cet effet des pseudonymes ou noms de plumes. Le philosophe français François Marie Arouet plus connu sous le nom de Voltaire a eu près de 173 pseudonymes (Docteur Atakia, Un Quaker, Zapata). Alors que madame Aurore Dudevant publiait sous un nom d'homme (George Sand), monsieur Charles Hamilton utilisait plutôt un nom de femme (Hilda Richards) pour ses publications. Quant aux chanteurs, Bob Dylan n'est nul autre que Robert Zimmerman, Cat Stevens reconverti en Yusef Islam est bien Stephen Demetri Georgiou; Tony Bennett, Anthony Benedetto (Nuessel 1992 : 19-23). Nombreux sont aussi les artistes qui requièrent l'anonymat, une autre façon de protéger leur identité.

Changements occasionnels : chez les Nandi, quand les hommes vont à la guerre, à la maison on ne doit plus prononcer leurs noms, on doit les appeler *oiseaux*.

Changements de nom pour une nouvelle vie : certains Inuit changent de noms quand ils deviennent très vieux dans l'espoir d'obtenir une nouvelle vie.

Changements de noms pour ne pas l'user : les Navahos, des Amérindiens, ont plusieurs noms, dont certains pour la guerre, car pour eux le nom contient un pouvoir et quand on l'utilise trop, il s'use. Il faut souvent le changer. Plus il est nouveau, plus il donne du pouvoir.

Changements de noms pour la protection : des ethnies en Afrique et ailleurs croyant que par le nom on peut jeter un sort à quelqu'un cachent leurs noms à des gens qui ne sont pas de leur famille. Ils se font connaître plutôt sous des sobriquets. Dans certaines ethnies de Madagascar, le nom d'un bébé est tabou car si on le prononce, on risque d'attirer sur le bébé le regard malveillant des ancêtres et ces derniers vont arracher l'enfant à la vie. Par conséquent, après avoir donné un nom de naissance à un bébé, on doit le cacher le plus possible et dissimuler la vraie identité de l'enfant par un second nom désagréable, répugnant, renvoyant à quelque chose qui repousse : Poubelle, Tas d'Ordures, Nain, Face de Chien, Crottin, Pourriture, Honte, etc. et on lui donne un troisième nom (un prénom occidental) pour l'école. Les adultes changent de nom chaque fois qu'un malheur les frappe. Autrefois, ils pouvaient le faire jusqu'à sept fois au cours de leur vie. Maintenant, une loi nationale limite le nombre de changement de noms à trois (Adler 1978 : 102). Dans certaines ethnies, quand quelqu'un meurt, les gens qui lui sont proches changent de nom de

peur que son esprit n'entende des noms familiers et revienne à la recherche de sa maison. Ailleurs, quand un enfant est malade, on change son nom pour détourner les mauvais esprits.

Également pour se protéger des ennemis, des personnes recherchées pendant ou après une guerre changent d'identité. La Deuxième Guerre mondiale, par exemple, a engendré beaucoup de changement de noms aussi bien parmi les Juifs que parmi les Allemands.

Changements de noms pour retrouver son identité :

Au sein de la population noire des États-Unis, il y a eu un grand courant de changement de nom pour prendre des noms purement africains. Cela a culminé au changement de l'appellation collective des Noirs : la dénomination Africains Américains remplace désormais Noirs Américains. Tout comme au Zaïre en 1972, sous Mobutu, le but visé est de préserver son identité en choisissant un nom qui sied au groupe en remplacement d'une étiquette inappropriée. Il n'y a pas longtemps, par exemple, les Inuit ont rejeté le nom *Esquimaux*, transposition française de «Eskimavak», mot d'origine algonquienne signifiant «mangeurs de viande crue», appellation péjorative. Ils ont préféré le terme qu'ils ont toujours utilisé eux-mêmes, à savoir «Inuit» signifiant «être humain, homme par excellence» (Hiu 1990 : 2). Les Innu ont eu la même réaction au sujet des appellations qu'on leur a données. Cette réaction, qui nous a inspiré pour le titre du présent article, et qui est clairement exprimée ici par un Innu, est fort éloquente :

«Le premier nom que le Blanc nous donna décrivait un animal, ou plutôt un cheval. Ce nom était "sauvage". On nous appela également "Indiens" et ce nom est encore utilisé aujourd'hui. Quant à moi, je n'aime pas qu'on m'appelle "sauvage" ou "indien". Maintenant depuis que nous avons été baptisés, on nous désigne sous le nom de Montagnais [...] et depuis quelque temps, il y en a d'autres qui nous appellent "Amérindiens". Ils nous donnent tous ces noms selon leur volonté, sans ce soucier de notre vrai nom qui est "Innu" [William-Mathieu Mark, Innu, 1993]» (Vitart 1995 : 8).

Ces paroles nous rappellent l'histoire de Kounta Kinté, cet esclave noir à qui, à son arrivée aux États-Unis, son maître voulait imposer le nom de Toby. Kounta Kinté cria partout, en réplique aux coups de fouet censés le réduire à accepter le nom Toby, qu'il s'appelait Kounta Kinté. Il fut battu presque à mort, mais refusa de renoncer à son nom (Haley 1973 : 172).

3.3 Quelques procédés de changement de nom

Comme le fait ressortir Albert Dauzat (1977 : 345), les changements de noms sont aussi différents que nombreux :

- **Substitution et modifications de forme** : simple modification d'orthographe (ou de phonétique) : de *Caumont* à *Chaumont* (pour franciser la physionomie d'un nom à consonance picarde).

- **Substitution d'un nom malsonnant ou ridicule** : (Badin, Couillard, Pineau, Bellegueule) ou simplement roturier, par un nom d'un ordre plus relevé (de Rochefort, de Vaucelles).

- **Substitution** des noms qui prêtent au calembour ou sont compris à contresens : *Bonabeau* (bon Abel, compris beau nabot).

- **Substitution d'un nom qu'une mauvaise orthographe rend déplaisant** : un dénommé *Mort* change de nom ignorant que c'était une mauvaise orthographe de *Mor*, ancienne forme populaire de *Maur*.

- **Substitution d'un nom à cause du sens péjoratif qui s'est développé** : *Garce* (fém. de *gars*), *Bordel* (ancien dérivé de *Borde*, ferme).

- **Substitution d'un nom s'il a été porté par un personnage fâcheusement connu** : *Juda*, rappelant *Judas Iscariote*; *Landru*, le tueur de femmes, exécuté en 1918 (ses enfants changèrent de nom) (Dauzat 1977 : 347).

Les noms étrangers n'échappent pas aux changements, surtout ceux qui sont difficiles à prononcer pour les nationaux. En France, pour s'assimiler au milieu, échapper aux railleries ou aux quolibets, tout en conservant certains éléments de leurs noms, des immigrants modifient leurs noms par diverses formes de combinaisons :

- **Amputation du nom par la finale ou par l'initiale** : *Feibelman* est raccourci en *Belmann* et *Van Ormelingen*, en *Vanor*.

- **Inversion des syllabes du nom** : *Cocu* devient *Cuco*.

- **Modification d'une ou deux lettres** : *Lévy* devient *Léry*, *Bloch* devient *Blot*.

- **Modification pour donner au nom une consonance voisine de celle du français** : *Deninger* devient *Denière*.

- **Remplacement d'un prénom ou d'un nom étrangers par ceux qui leur sont proches et qui sont connus dans le nouveau pays** : *Dedominici* devient *Dominique*, *Butscher* devient *Boucher* (Dauzat 1977 : 349).

- **Traduction des noms étrangers dans la langue du nouveau pays** : *White* devient *Blanc* (Dauzat 1977 : 350).

- **Nom de l'épouse** : si l'étranger a épousé une Française, il prend le nom de sa femme, ou l'adjoint au sien, ou ne donne aux enfants que le nom de leur mère (Dauzat 1977 : 360).

Au Zaïre, à la suite de la décision de 1972 qui stipulait que le nom devait être trouvé exclusivement dans le patrimoine culturel zaïrois, certains Zaïrois ont utilisé des procédés assez subtils pour changer les noms :

a) Chez les Bakóngo :

- **Transformations de graphie et de prononciation des noms européens** : Salóóti, de Charlotte; Maádi, de Marie; Lofóoso, de Alphonse; Tééya, de Théophile;

- **Adaptation phonétique et morphologique des noms européens** : Bols est devenu Bola.

b) Chez les Bálúba :

- **Substitution d'anthroponymes étrangers par des noms lúba ne présentant aucune ressemblance linguistique mais attribués d'après une caractéristique morale ou physique de ceux qui les reçoivent** : Cyélanténdá, 'Lanceur des flèches' : Père Omar Boghemans; Dfíbalá, 'Calvitie' : Commandant Pelzer; Kákúngúlá, 'Le Rasé' : Frère Hippolyte Fiers; Lúkombó, 'Le Balai' : Père Van Landschoot; Mpándámábwé, 'Tailleur de pierres' : Frère Jan Damiaens; Talatála, 'Lunettes' : Père Jules Garmyn. Les porteurs de ces noms les acceptaient et s'identifiaient eux-mêmes sous ces appellations (Madiya 1991 : 55).

L'identité personnelle et l'héritage culturel nous amènent à dire un mot sur la traduction des anthroponymes. Le contenu culturel et identitaire des anthroponymes est si important et si personnel qu'on se rend à l'évidence des paroles de John Lyons (1978 : 181) : «Même s'il existe un équivalent de traduction reconnu, il n'est pas toujours à propos de l'utiliser». Il faudrait, cependant, nuancer, car la finalité de l'heure et la situation dans laquelle on se trouve peuvent, de manière exceptionnelle, justifier l'utilisation d'un équivalent.

RÉFÉRENCES

ADLER, Max Kurt (1978) : *Naming and Adressing : a Sociolinguistic Study*, Hamburg, Helmut Buske Verlag Hamburg, 282 pages.

ALFORD, Richard D. (1988) : *Naming and Identity : a Cross-Cultural Study of Personal Naming Practices*, New Heaven Connecticut, HRAF Press, 190 pages.

BAL, Willy (1991) : «Limites de la présente monographie», Bal, Willy, Jan Daeleman et Clémentine Faïk-Nzuji Madiya, *Anthroponymie afro-romane, esquisse d'un projet*, Tübingen, Max Niemeyer Verlag, pp. 3-5.

Britannica (1991) : *Encyclopaedia Britannica*, vol. 24.

DAUZAT, Albert (1977) : *Les noms de famille de France; traité d'anthroponymie française*, 3^e édition révisée et complétée par M.T. Morlet, Paris, Librairie Guenegand, 371 pages.

HALEY, Alex (1977) : *Racines*, traduit de l'américain par Maud Sissung, [Lieu de publication non indiqué], Éditions Alta, 476 pages.

HIU, Lucile (1990) : *Les Esquimaux de Sibérie révélés par leurs contes : mythes et quotidien*, traduction et adaptation, Paris, Jean Maisonneuve Éditeur-Libraire, 269 pages.

- ILBOUDO, T. J. (1977) : «Réflexions sur le nom dans la tradition mossi et perspectives chrétiennes», Ceeba Publications (dir), *Noms théophores d'Afrique*, Série 11, Vol. 45, Bandundu, République du Zaïre, pp. 105-111.
- LEFEBVRE-TEILLARD, Anne (1990) : *Le nom : droit et pouvoir*, 1^{re} édition, Paris, Presses Universitaires de France, coll. «Léviathan», 247 pages.
- LIVI BACCI, Massimo et Lorenzo DEL PANTANELL (1974) : «Identification des individus à partir du XVII^e siècle en Italie», Louis Henry (dir), *Noms et prénoms : aperçu historique sur la dénomination des personnes en divers pays*, Dolhain, Ordina Éditions, pp. 83-98.
- LYONS, John (1978) : *Éléments de sémantique*, traduit par Jacques Durand avec la collaboration d'Éliane Koskas, Paris, Librairie Larousse, coll. «Langue et Langage», 296 pages.
- MADIYA, Clémentine Faïk-Nzuji (1991) : «Ce que nom propre signifie», Bal, Willy, Jan Daeleman et Clémentine Faïk-Nzuji Madiya, *Anthroponymie afro-romane, esquisse d'un projet*, Tübingen, Max Niemeyer Verlag, pp. 11-14.
- MADIYA, C. Faïk-Nzuji (1991) : «Noms d'origine étrangère dans l'anthroponymie zairoise : le cas des noms luba», Bal, Willy, Jan Daeleman et Clémentine Faïk-Nzuji Madiya, *Anthroponymie afro-romane, esquisse d'un projet*, Tübingen, Max Niemeyer Verlag, pp. 47-56.
- MADIYA, Clémentine Faïk-Nzuji (1991) : «Système anthroponymique lúba», Bal, Willy, Jan Daeleman et Clémentine Faïk-Nzuji Madiya, *Anthroponymie afro-romane, esquisse d'un projet*, Tübingen, Max Niemeyer Verlag, pp. 15-25.
- NELL, Adelheid von (1974) : «Les noms de famille et les prénoms en Allemagne», Louis Henry (dir), *Noms et prénoms : aperçu historique sur la dénomination des personnes en divers pays*, Dolhain, Ordina Éditions, pp. 5-14.
- NISSIM, G. M. (1977) : «Les noms théophores chez les Bamiléké de Jo», Ceeba Publications (dir), *Noms théophores d'Afrique*, Série 11, Vol. 45, Bandundu, République du Zaïre, pp. 56-66.
- NUESSEL, Frank (1992) : *The Study of Names : A Guide to the Principles and Topics*, Westport, Greenwood Press, 153 pages.
- RETEL-LAURENTIN, A. et S. HORVATH (1972) : *Les noms de naissance (indicateurs de la situation familiale et sociale en Afrique Noire)*, Paris, Centre national de la recherche scientifique, coll. «Société d'études linguistiques et anthropologiques de France», 160 p.
- SANCHEZ-ALBORNOZ, Nicolas (1974) : «La dénomination des personnes en Amérique Espagnole», Louis Henry (dir), *Noms et prénoms : aperçu historique sur la dénomination des personnes en divers pays*, Dolhain, Ordina Éditions, pp. 15-20.
- VITART, Anne (1995) : *Les Indiens montagnais du Québec entre deux mondes*, Paris, Éditions Sépia, 125 pages.

LES VARIANTES GRAPHIQUES ET LEUR SOURCE DE VARIATION

André DUGAS

Université du Québec à Montréal, Canada

INTRODUCTION

L'étude du français révèle un nombre relativement important de variantes graphiques. On reste quelque peu interdit devant ce fait, puisque nous savons que la variation sur le même mot, qui peut aller à l'infini, est contrée par les lexicographes dont l'une des tâches est de préserver l'«unicité» de la représentation du stock lexical dans les ouvrages dictionnaires. Il en résulte, par exemple, que les mots *tomate* ou *vérité* n'ont pas de graphie autre. Il se produit cependant que plusieurs graphies soient proposées pour un seul mot. L'une des explications les plus raisonnables d'une graphie hésitante peut venir du fait que la langue transcrite ne jouit pas d'une longue tradition littéraire, ou tout simplement graphique, ou qu'aucun organisme régulateur n'est voué expressément à la standardisation de la graphie. Bien entendu, ce n'est pas la situation du français. Se pourrait-il que l'hésitation entre deux graphies s'explique de façon paradoxale par les errances dans le passé de mots qui ont traversé les siècles de façon différente ? C'est en voulant répondre à cette question que cette étude a été conduite.

La variation graphique n'a jamais fait l'objet et ne fera probablement jamais l'objet d'une explication rationnelle. À première vue, on peut cependant distinguer les variantes de formes sans changement phonétique de celles qui entraînent un tel changement; pour les premières, on observe *déclancher* parallèlement à *déclencher*, pour les secondes, *écarbouiller* et *écrabouiller*. On observe également des variantes qui s'appuient sur une alternance morphologique, faisant varier les affixes verbaux, côté préfixe ou côté suffixe, par exemple, *affraîchir* et *fraîchir*, ou *jazzier* et *jazzifier*.

On se rend compte enfin que des variantes diffèrent autrement, attestant d'un cheminement historique particulier (ces variantes sont souvent étiquetées *synonymes*; les créations argotiques de ce type foisonnent), par exemple, *abécher* et *abecquer*.

On peut supposer que la plus grande partie de l'alternance correspond à des formes de la première sorte et que, dans la mouvance des différentes tentatives de réforme de l'orthographe, c'est la réduction de ces formes qui est visée. L'étude effectuée, comme nous le verrons, avait aussi pour but de vérifier si les alternances sont sujettes à réduction dans le cas où les responsables d'une réforme de l'orthographe, fortuite et, maintenant tout aussi improbable, pourrait tenir compte de nos suggestions et d'autres de ce genre.

L'étude a porté sur une liste de 10 000 verbes dont 317 d'entre eux révèlent au moins une forme en variation, ce qui veut dire plus de 650 verbes différents. Tous les verbes étudiés sont attestés dans l'un ou l'autre des dictionnaires mentionnés dans la bibliographie¹.

DESCRIPTION DES SOUS-CLASSES DE VARIANTES

Les classes de variantes se départagent de la manière suivante; les formes en variation peuvent influencer la prononciation ou être sans effet. Voyons d'abord les sous-classes de variantes verbales sans changement phonétique.

Variantes sans changement phonétique

1° C > CC (62 entrées², soit 19,6 %)

On le sait, l'orthographe du français est aberrante, et c'est bien l'une des rares langues où deux consonnes graphiques consécutives ne sont nullement l'indication d'une longueur phonétique. C'est pour cette raison que nous commencerons notre examen par l'exploration des verbes détenteurs d'une double consonne en variation. Il nous vient aussi à l'esprit que le redoublement de consonnes est l'un des facteurs courants de la variation graphique : un verbe sur cinq de notre liste est en effet affecté par ce phénomène. Cette liste contient, par exemple, *beloter/belotter*, *frisoter/frisotter*, *grafigner/graffigner*, *racastiller/raccastiller*, *snifer/sniffer*. Si l'on observe encore les variantes *boursoufler/boursouffler*, *embatre/embattre*, on vérifie cependant que *battre* et *souffler* n'ont pas de variante avec un seul f; de la même manière; s'il y a *consoner/consonner*, *gafer/gaffer*, on constate que les variantes nominales *consonne* (pour *consonne*), *gafe* (pour *gaffe*) ne sont pas attestées.

La variation affecte la moitié de ces verbes dans leur radical, comme *afistoler/affistoler* ou *frottailler/frotttailler*³. Pour les autres verbes, ce sont les consonnes qui précèdent immédiatement la terminaison qui sont redoublées. Lors de la dernière tentative de réforme de l'orthographe, les verbes en *-eler* et *-eter* ont fait l'objet d'une attention particulière dans le document du Conseil supérieur de la langue française (p. 14) : «Il existe deux procédés pour noter le E ouvert :

- soit le redoublement de la consonne qui suit l'E (exemple : *ruisselle*);

- soit le E accent grave, suivi d'une consonne simple (exemple : *harcèle*).»

Les verbes en *-eler* n'ont cependant pas de variante à l'infinitif dans notre liste, et il n'y en a qu'une dizaine pour les verbes en *-eter* (*caneter/canetter*, *levreter/levretter*, etc.)⁴.

¹ Une bonne partie de cette étude a été effectuée dans le cadre de travaux exécutés au Laboratoire de linguistique informatique de l'Université Paris 13, placé sous la direction du Professeur Gaston Gross. Je dois beaucoup à M. Michel Mathieu-Colas qui y conduit ses propres recherches sur le verbe du français; il m'a fait profiter de ses travaux sur le sujet, de ses tables, de ses tris, sans compter son temps, ses commentaires et ses conseils.

² Il s'agit de l'entrée de l'un des verbes en variation, sur 317 entrées.

³ Dans cet exemple, il faut cependant considérer que le suffixe est *-ailler* et non simplement *-er*.

⁴ Cela ne veut pas dire que les spécialistes appelés à faire des propositions de réforme ont mal vu. On peut cependant supposer qu'en désignant les seuls verbes en *-eler* et *-eter*, ils ne

En revanche, d'autres verbes, qui ne sont pas soumis à la contrainte du E ouvert, sont susceptibles de révéler une variante à double consonne avant terminaison, dont une quinzaine de verbes en -oter : *bécoter/bécotter*⁵, *frisoter/frisotter*, *mangeoter/mangeotter*, etc. Des variantes de la sorte sont attestées pour des verbes dont la finale est en -aler, -aner, -aper, -ater, -atre, -ifer, -ofer, -olir, -oner, -oufer, -oufler, -ouiser, -ourer.

2° h <> Ø (4 entrées, soit 1,3 %)

La présence ou l'absence de la consonne h n'a pas de conséquence pour la prononciation. La variation concerne peu de verbes. Dans nos listes, il n'y en a que quatre; ce sont *empapahouter/empapaouter*, *enchrister/encrister*, *hyperhémier/hyperémier*, *rhapsoder/rapsoder*.

3° C₁ <> C₂ (19 entrées, soit 6 %)

Il s'agit de groupes de consonnes dont la prononciation est équivalente dans un contexte donné. Sont rangées dans cette catégorie plusieurs variantes dont voici des exemples :

c/ck; tacler/tackler

c/qu; coailler/quoailler, picouser/piquouser, taconner/taquonner

c/que; décocter/décoqueter

c/cque; débecter/débecqueter, jacter/jacqueter

ch/sch; chpiler/schpiler, chnoufer/schnouffer

g/j; galéger/galéjer

s/es/esse; smiller/esmiller/essemiller

s/ce; balanstiquer/balancetiquer, lansquiner/lancequiner, lanscailler/lancecailler

s/sse; lisbroquer/lissebroquer

s/z; ébouser/ébouzer, piquouser/piquouzer

v/w; envagonner/enwaggonner

Voilà pour les consonnes. Des voyelles sont également impliquées dans les phénomènes de variance sans changement phonétique. La variation graphique concerne plusieurs voyelles.

4° Les variantes vocaliques : an <> en, am <> em, au <> o, ai <> è, oua <> a, au <> ô, ai <> ei, ou <> oo, oa <> o (20 entrées, soit 6,3 %)

disposaient pas de listes suffisamment explicites pour mesurer l'ampleur du phénomène du redoublement de la consonne et de l'hésitation qui marque bien davantage de mots du français qu'ils n'ont pu alors le supposer.

⁵ Pour l'emploi de *bécoter*, avec un seul t, on cite Brassens (Le Grand Robert). Le lexicographe, en citant le seul auteur de la graphie particulière d'un mot, tente sans doute par là d'en restreindre l'«étendue d'emploi» — ce faisant, il ne fait que respecter une pratique commune aux grands dictionnaires. Le doute demeure : un auteur a-t-il délibérément ou non travesti un mot, a-t-il fait une faute d'orthographe d'usage à son insu, était-il simplement hésitant sur l'orthographe d'un mot dont on lui attribue maintenant la paternité ?

Il n'y a que les alternances an/en et au/o qui présentent plus d'un candidat : *antifler/entifler, calancher/calencher, déclancher/déclencher, lanter/lenter, ranquiller/renquiller*, puis *auloffer/oloffer, dépiauter/dépioter, endauffer/endoffer, flauper/floper, rabiauter/rabioter*. La variation am/em est réservée à *étramper/étremper*. Il en est de même des autres variantes : *empouacrer/empoicrer, entauler/entôler, faignanter/feignanter, shouter/shooter, toaster/toster*. Une mini-série présente des anomalies : il y a *dégrainer/dégréner*, mais *dégrener* est attesté, de même que *engrainer/engrener* et *graineter/grener*.

Peu de variantes enregistrées dépendent d'alternances graphiques du genre *craspeler/crêpeler, décraspouiller/décrapouiller*.

Des glissantes participent également à la variation, par exemple, *bordayer/bordailier, faséyer/faseiller/faséier/faseyer*.

5° e <> Ø (4 entrées, soit 1,3 %)

Comme pour le h muet, le schwa peut être éliidé. Voici les exemples relevés dans notre corpus : *briftonner/brifetonner, calcer/calecer, chlinguer/chelinguer, enchtiber/enchetiber*.

6° - <> Ø (22 entrées, soit 6,9 %)

Le trait d'union est une marque purement graphique qu'on trouve surtout dans les unités composées. Dans notre liste, il caractérise la «composition» de trois espèces de verbes. Il y a la série avec entre- — il s'agit de 14 entrées — dont *entre-déchirer/entredéchirer*; parfois, le trait d'union est remplacé par une apostrophe après élision du schwa comme dans *entre-égorger/entr'égorger*, ou enfin, la variation ne correspond qu'aux deux variantes du type suivant, *entr'ouvrir/entrouvrir*. Une autre petite série de 4 entrées est composée de *ré-écouter/réécouter, ré-écrire/réécrire, ré-imposer/réimposer, ré-orienter/réorienter*. Une dernière liste disparate est composée des unités *arc-bouter/arcbouter, auto-mutiler/automutiler*⁶, *bi-socier/bisocier, don-juaniser/donjuaniser*.

Variantes avec changement phonétique

Dans ce qui suit, nous explorons les sous-classes dont les variantes attestent un changement phonétique.

7° Suffixes de conjugaison er <> ir <> oir <> (r)e (18 entrées, soit 5,7 %)

Des variantes de verbes sont partagées entre plusieurs types de conjugaison. L'étude des verbes de cette sous-classe nous fait plonger dans une diachronie «profonde», et il n'y a

⁶ Le préfixe auto- fait partie des préfixes à la mode en néologie : le nombre d'unités attestées, toutes de création récente, dépasse largement la centaine; il a été encore possible de bien former quelque 800 néologismes virtuels (Dugas, 1992 : 24-29). De façon non attestée, les variantes pullulent, avec ou sans trait d'union.

pas de surprise à voir des membres de ces entrées caractérisés par les qualificatifs «rare» ou «vieux». Voici les exemples relevés dans notre liste :

a. er/ir

ablater/ablatir
alester/alestir
alloter/allotir
amocher/amochir
forhuer/forhuir

grincher/grinchir
mater/matir
rabouter/raboutir
rengracier/rengracir
sailler/saillir

b. er/oir

arder/ardoir

permaner/permanoir

c. er/(r)e

arder/ardre
attirer/attraire
bruïsser/bruïre

ficher/fiche
tisser/tistre, tître, titre
rentrayer/rentraire

8° suff₁ <> suff₂ (34 entrées, soit 10,7 %)

Un second groupe de suffixes caractérise le verbe. En voici des exemples, de plusieurs variétés.

argoter/argotiser, chloroformer/chloroformiser, subjectiver/subjectiviser, voyeller/voyelliser;
agatiser/agatifier, baroniser/baronifier, élégatiser/élégantifier, jazer/jazzifier;
emmurer/emmurailler, licher/lichailler;
causailler/causoter/causotter;
boitailler/boitiller;
fumailler/fumasser;
brandiller/brandouiller;
ensiler/ensiloter;
flânoter/flanocher, pleuvioter, pleuvocher;
mater/matouser, valiser/valouser;
démailler/démaillonner, rejeter/rejetonner;
dégraver/dégraveler.

9° préf₁ <> préf₂ (29 entrées, soit 9,1 %)

Des variantes sont également attestées, qui dépendent de la préfixation verbale. En voici des exemples :

affraîchir/fraîchir, aguetter/guetter;
revendiquer/vendiquer;
désamboûter/déboûter, déséchouer/déchouer, désaffubler/défubler;
délaiter/élaiter.

Une série nous vient du maintien ou non du é de ré-. Voyons des exemples.

réaccoutumer/raccoutumer, réélargir/rélargir, réensemencer/ensemencer.

10° rad₁ ⇔ rad₂ (105 entrées, soit 33,1 %)

La description de cette sous-classe demande la plus grande précision. On peut auparavant se poser la question suivante : comment est-elle homogène ? Faisons ensemble une révision : la variation qu'on observe pour ses membres ne peut plus affecter ses affixes, ce que nous avons vu plus haut. Il s'agit donc d'examiner de plus près le radical du mot et de noter des différences phonétiques le plus souvent minimales. Ces variantes sont des témoins uniques de la difficulté de n'enregistrer qu'une graphie. Le fait de prononcer différemment n'est pas exceptionnel; ce qui l'est, c'est de retenir les produits de la divergence, autrement dit de noter officiellement les différents mots témoins d'une hésitation. En guise d'entrée en matière, voyons des exemples de plusieurs variantes argotiques synonymes du verbe *mourir*, de deux verbes empruntés à une langue étrangère et de deux termes techniques :

clamser/clamecer/cramser/crampser/crapser/clampser
jodler/yodle/iodler/iouler
poter/putter
enjaler/enjauler/enjouailler/enjoualer
plamer/plaimer/pelaner/planer.

Dans tous les cas, l'une ou l'autre graphie correspondent à un emploi davantage populaire, largement diffusé localement, du moins suffisamment pour que le lexicographe hésite à n'inscrire que l'une ou l'autre des variantes. Voyons encore des exemples plus courants : la description des deux séries est satisfaite par l'un ou l'autre des patrons expliqués ci-après :

acupuncturer/acuponcturer
conteneuriser/containériser/containeriser
converser/conversationner, jacter/jacquetancer, maller/mallouser
déraïdir/déroidir
léchotter/lichotter
médiciner/médeciner,
pasteuriser/pastoriser
pioncer/pioncetiquer
poigner/pogner, encoigner/encogner, magner/manier
revenger/revancher
substanter/sustenter
tarifer/tarifier.

Les variantes dues au processus de la métathèse sont de la même veine, par exemple :

bertauder/bretauder//bertouder/bretouder
éberner/ébrenner
écarbouiller/écrabouiller

ou de l'alternance l/r,

connobler/conobler/connobrer/conobrer.

D'autres variantes dépendent de l'alternance d'une consonne sourde et d'une sonore, d'une nasale et d'une orale, au même point d'articulation ou non, d'une occlusive et d'une continue, etc., d'une voyelle plus ou moins ouverte, d'une nasale et d'une orale, etc. Voyons des exemples :

(C₁ <> C₂)

débarouler/dégarouler

(C sonore <> C sourde)

fauberder/fauberter, terreauder/terreauder, palanguer/palanquer, chacler/châbler

(C orale <> C nasale)

flamber/flammer, pessiller/pessigner, tromboller/tromboner

(C nasale₁ <> C nasale₂)

argouiner/argouigner

(C occlusive <> C continue)

abécher/abecquer, ségréger/ségréguer

(V orale <> V nasale)

rabaner/rabanter

(V ± haute)

agrapper/agripper

passager/passéger, manualiser/manuéliser

abréyer/abrier

guaper/gouêper

bécher/bêcher, bléser/blaiser

boïtter/bouetter

repamer/repaimer, surjaler/surjauler

rocouer/roucouer, cronir/crônir/crounir

ébotter/ébouter, portraire/pourtraire

tromboner/trombiner

affeurer/afforer

aveuer/avuer

embreuver/embrever

engrainer/engrener, haleiner/halener

(V haute₁ <> V haute₂)

biduler/bidouiller

(V e ⇔ é)
querir/quérir, retreindre/rétreindre, reventer/réventer, reviser/réviser, schueller/schuéler.

D'autres variantes ne se distinguent que par l'ajout — ou la suppression — d'une consonne (ou une semi-consonne), comme dans :

abrier/abriter, affurer/afflurer, ajuster/ajuster, bivaquer/bivouaquer, clairer/claircer, coucouer/coucouler, morfier/morfigner/morfiller, saper/sarper, chapuser/chapuser.

Des variantes se distinguent également par l'ajout — ou la suppression — d'une voyelle :

chosifier/choséfifier.

Des variantes résultent d'une amalgame des opérations décrites ci-dessus, par exemple :

aplaigner/aplaner, chromer/chromater, clatir/glapir, coqueliner/coqueriquer, coupeter/copter, énoper/énouer, fiasquer/flasquer, gruer/gruauter, mostaurder/moutarder, panader/pavaner, portraiturer/portraicturer/pourtraicturer, rouster/roster/rousturer, tapouiller/tapoter/tapouser, tripolir/tripolisser.

La description d'un dernier groupe de variantes paraît bien aléatoire; qu'on en juge (il s'agit de la liste complète) :

choséfifier/matérialiser/réfifier, corailler/croasser, dérâper/égrapper, florer/espalmer, panoter/panoramiquer, remplir/rendoubler, ringardiser/ringuer, tagger/taguer, tireler/tirelirer, tôler/gondoler.

CONCLUSIONS

Nous venons de voir un système complet de variantes d'un sous-ensemble du lexique du français. En guise de conclusion, voyons les commentaires qui s'imposent. Ils sont de deux ordres. On constate d'abord, bien que l'échantillon décrit ne soit pas très important, que la variation graphique n'est pas le fouillis que l'on pourrait imaginer. Les schémas décrits s'appuient sur des phénomènes oraux depuis longtemps observés pour des milliers de mots pris individuellement : quel instituteur n'a pas repris un élève pour son interversion dans *aéroplane* ? Qui n'abuse pas de *classifier* aux dépens de *classer* ? La plupart des verbes en variation (à presque 60 %) appartiennent à ce groupe, des schémas 7° à 10° dans le Tableau récapitulatif.

Notre attention est ensuite attirée par les variantes sans répercussion à l'oral, c'est-à-dire celles qui ne comportent pas de changement phonétique — qui n'a pas oublié un c dans *occurrence* ? Elles sont décrites dans les schémas de 1° à 6° inclusivement. Elles sont directement rattachées à l'hésitation dans l'écriture. Ce 40 % de variantes devrait être un terrain privilégié des spéculations que ravivent les tentatives successives de réforme de l'orthographe.

**Tableau récapitulatif
Sous-classes des variantes
des verbes du français**

1° C <> CC (62 entrées, soit 19,6 %)	<i>beloter/belotter, frisoter/frisotter, grafigner/graffigner, racastiller/raccastiller, snifer/sniffer * boursoufler/boursouffler, embatre/embattre ** consoner/consonner, gafer/gaffer</i>
2° h <> Ø (4 entrées, soit 1,3 %)	<i>encrister/enchrister</i>
3° C ₁ <> C ₂ (19 entrées, soit 6 %)	<i>tacler/tackler</i>
4° Les variantes vocaliques : an <> en, am <> em, au <> o, ai <> è, oua <> a, au <> ô, ai <> ei, ou <> oo, oa <> o (20 entrées, soit 6,3 %)	<i>antifler/entifler shouter/shooter rabiauter/rabioter dégrainer/dégrèner</i>
5° e <> Ø (4 entrées, soit 1,3 %)	<i>calcer/calecer</i>
6° - <> Ø (22 entrées, soit 6,9 %)	<i>entre-déchirer/entredéchirer, arc- bouter/arcbouter</i>
7° Suffixes de conjugaison er <> ir <> oir <> (r)e (18 entrées, soit 5,7 %)	<i>ablater/ablatir, arder/ardoir, bruissier/bruire</i>
8° suff ₁ <> suff ₂ (34 entrées, soit 10,7 %)	<i>argoter/argotiser, ensiler/ensiloter démailler/démaillonner</i>
9° préf ₁ <> préf ₂ (29 entrées, soit 9,1 %)	<i>affraîchir/fraîchir, désemboîter/déboîter, jazzier/jazzifier</i>
10° rad ₁ <> rad ₂ (105 entrées, soit 33,1 %)	<i>débarouler/dégarouler, argouiner/argouigner, abécher/abecquer, agrappier/agripper</i>

RÉFÉRENCES

CATACH, N., GOLFAND, J. et R. DENUX (1971) : *Orthographe et lexicographie* (Littré, Robert, Larousse) Tome I *Variantes graphiques - Mots latins et grecs - Mots étrangers*, Publications du Centre d'étude du français contemporain, sous la direction de B. Quemada, vol. 3, Paris, Didier.

Conseil supérieur de la langue française (1990) : *Rapport du groupe de travail sur Les rectifications de l'orthographe*, Paris, Hôtel de Matignon.

DUGAS, A. et H. MANSEAU (1996) : *Les verbes logiques*, Montréal, Les Éditions Logiques.

DUGAS, A. (1995) : «Variation orthographique», *Équivalences*, Revue de l'Institut supérieur de traducteurs et interprètes de la Communauté française de Belgique, vol. 24/2 et 25/1-2, 1994-1995.

DUGAS, A. (1992) : «Le préfixe auto-», A. Dugas et Ch. Molinier (dir), *La productivité lexicale*, dans *Langue française*, n° 96, Paris, Larousse, pp. 20-29.

DUGAS, A. (1990) : «La création lexicale et les dictionnaires électroniques», Bl. Courtois et M. Silberztein (dir), *Dictionnaires électroniques du français*, dans *Langue française*, n° 87, Paris, Larousse, pp. 23-30.

MATHIEU-COLAS, M. (1990) : «Orthographe et informatique : établissement d'un dictionnaire électronique des variantes graphiques», Bl. Courtois et M. Silberztein (dir), *Dictionnaires électroniques du français*, dans *Langue française*, n° 87, Paris, Larousse, p. 104-111.

REY-DEBOVE, J. et B. LE BEAU-BENSA (1991) : *La réforme de l'orthographe au banc d'essai du Robert*, Paris, Dictionnaires Le Robert.

THIELE, J. (1987) : *La formation des mots en français moderne*, Presses de l'Université de Montréal, Traduction de A. Clas de *Wortbildung des französischen Gegenwartssprache*, Leipzig, Verlag Enzyklopädie, 1985.

Dictionnaires dépouillés

COLIN, J.-P., MÉVEL, J.-P. et C. LECLÈRE (1990) : *Dictionnaire de l'argot*, Paris, Larousse.

Grand dictionnaire encyclopédique Larousse (1982-1985) : 10 vol., 1 supplément (1992), Paris, Librairie Larousse.

Grand Larousse encyclopédique (1960-1964) : 10 vol., 2 suppléments, Paris, Librairie Larousse.

HATZFELD, A., DARMESTETER, A. et A. THOMAS (1932) : *Dictionnaire général de la langue française - du commencement du XVII^e siècle jusqu'à nos jours. Précédé d'un Traité de la formation de la langue*, 2 vol., Paris, Librairie Delagrave.

Larousse du XX^e siècle (1928-1933) : 6 vol., 1 supplément, Paris, Librairie Larousse.

Le Grand Robert de la langue française (1985) : Dictionnaire alphabétique et analogique de la langue française de Paul Robert, 2^e édition, entièrement revue et enrichie par Alain Rey, Paris, Le Robert.

Le Nouveau Petit Robert (1993) : *Dictionnaire de la langue française 1*, Paris, Dictionnaires Le Robert.

Nouveau Larousse illustré (1896-1904) : 7 vol., 1 supplément, Paris, Librairie Larousse.

ROBERT, P. (1966) : *Dictionnaire alphabétique et analogique de la langue française - Les mots et les associations d'idées*, 6 vol., 1 supplément (1976), Paris, Le Robert.

Trésor de la langue française (1985-1994) : 15 vol., Paris, CNRS et Klincksieck, puis Gallimard.

TEL MÉTIER TEL MOT : LES CONDUCTEURS DE TAXI-MOTO «ZÉMIDJAN» DE COTONOU¹

Séverin-Marie KINHOU

Centre National de Linguistique Appliquée, Cotonou, Bénin

INTRODUCTION

La décision de réaliser un plan décennal d'aménagement linguistique de la francophonie contenue dans la déclaration solennelle de Dakar² vise deux objectifs :

«organiser de façon harmonieuse et efficace, en particulier dans les États du Sud, la coexistence du français, langue commune et facteur d'unité de l'espace francophone, et des autres langues de ces États et de cet espace dont l'existence est reconnue, comme langues officielles et/ou nationales [...] Toutefois, cet acte a une portée infiniment plus grande, car il constitue aussi un élément déterminant du développement économique, social et humain des États du Sud.»

La gestion des plurilinguismes nationaux, se fondant sur une connaissance des situations sociolinguistiques spécifiques à chaque État, est aussi une condition de

¹ Quelques points de repères du Bénin extrait de *L'année francophone internationale*, ACCT et AFELSH (1997) : AFI, Québec, pp. 166-167.

Ancien royaume du Dahomey célèbre par ses rois Gléglé et son fils Béhanzin qui luttèrent contre la France jusqu'en 1894. Membre de l'ancien AOF (Afrique Occidentale Française), pépinière des cadres de l'Afrique, d'où son surnom de «Quartier latin de l'Afrique», il accède à l'indépendance en 1960. De 1972 à 1989, sous Mathieu Kérékou, le Dahomey devient République populaire du Bénin avec le marxisme-léninisme comme idéologie d'État. Battu aux élections présidentielles par Nicéphore Soglo en 1991 grâce au vent de la démocratie qui soufflait alors sur le continent, le Général Mathieu Kérékou retourne aux affaires en 1996 dans le cadre de l'actuelle constitution. Porto-Novo (180 000 habitants) est la capitale officielle mais sa voisine, Cotonou (550 000 habitants), est la métropole qui abrite le gouvernement.

Pays du Golfe de Guinée, le Bénin couvre une superficie de 112 622 km² avec une population estimée en 1997 à environ 5 000 000 d'habitants. Le Bénin compte 52 groupes socio-culturels dont les plus majoritaires au sud sont : le groupe Fon : 42,2 %, le groupe Adja : 15,6 % et le groupe Yoruba : 12,1 %; au nord, on y trouve le groupe Bariba : 8,6 %, le groupe Otamari : 6,2 %, et le groupe Dendi : 2,8 %.

² Le Sommet de Dakar d'abord et celui de Chaillot ensuite ont décidé la réalisation d'un plan d'aménagement décennal linguistique.

participation du plus grand nombre au processus de développement et ouvre ainsi la voie à la démocratie.

Mais, la multiplicité des actions à mener impose la nécessité de la mise en route des structures nationales d'aménagement linguistique, cadres officiels de gestion de la politique linguistique au niveau de chaque État.

L'État doit aménager les langues nationales de grande communication et commencer leur instrumentalisation fonctionnelle.

Avant de vous présenter mon exposé étymologique sur un terme relevant du secteur des transports qui est en perpétuelle mutation, permettez-moi de partager avec vous quelques concepts d'aménagement linguistique.

Généralités

Langue et développement sont étroitement liés du fait que l'éducation est indispensable pour la formation de la main d'œuvre spécialisée nécessaire à l'entreprise de développement national.

En effet, la langue est le moyen par lequel les savoirs : savoir-être, savoir-faire et les compétences sont dispensés dans le processus de l'éducation et dans la formation de la main d'œuvre. D'où l'efficacité du processus de formation et de l'éducation dépend souvent de l'efficacité de la communication du savoir-faire par le moyen de la langue.

La planification linguistique pour Eastman (1983), est motivée par le fait que la langue est perçue comme une ressource sociale et dès lors, doit être planifiée si l'on veut l'utiliser dans toutes ses potentialités. Cette planification ne peut s'opérer que dans un plan d'aménagement linguistique spécifique à chaque État.

En quoi consiste donc l'aménagement linguistique ?

C'est le processus éclairé et réfléchi par lequel sont conçus, élaborés et appliqués des programmes pour l'utilisation des langues dans une communauté linguistique ou dans un pays.

Il comprend généralement deux volets : l'aménagement du statut de la langue et l'aménagement du corpus de la langue elle-même.

L'aménagement du statut de la langue est une opération d'ordre politique qui consiste à conférer à une ou plusieurs langues un statut officiel, des fonctions précises et un cadre juridique en ce qui concerne son usage dans les communications internes et externes dans une communauté ou une institution donnée. Sur cette question du choix des langues, il existe encore un flou au niveau de l'article 11 de la nouvelle Constitution³ du Bénin où le français jouit du statut de langue officielle et d'enseignement.

³ Toutes les communautés composant la nation béninoise jouissent de la liberté d'utiliser leurs langues parlées et écrites et de développer leur propre culture tout en respectant celle des autres.

En tant qu'instrument de communication, la langue elle-même doit être constamment aménagée et s'adapter aux activités professionnelles de tous les secteurs de la vie et surtout dans les pays du sud où se développe le secteur dit «informel».

À profession nouvelle terminologie nouvelle, dit-on.

Il s'agit donc de doter nos langues de nouvelles terminologies nécessaires pour exprimer les nouvelles réalités.

Ce qui veut dire qu'il faut inventer, selon Clas (1985 : 57) : «les nouvelles "étiquettes" à mettre sur les concepts indispensables, sans lesquelles toute connaissance et toute structuration deviennent difficiles». Les termes créés par conséquent doivent répondre, entre autres, à des exigences de :

- conformité au système linguistique (structures morphophonologiques);
- conformité aux habitudes socio-culturelles de la communauté linguistique;
- conformité directe avec les activités socio-professionnelles, technologiques ou scientifiques du public utilisateur de la langue.

Clas (1985) distingue deux types de langue : la langue commune qui sert dans tous les usages de la même communauté linguistique et les langues de spécialités utilisées par des groupes déterminés dans des domaines spécifiques.

Pour illustrer nos propos, nous avons choisi une activité du secteur dit «informel» qui pose non seulement un problème terminologique mais d'autres chapelets de problèmes de développement tout court aussi bien aux autorités administratives et politiques qu'à ses usagers.

Nous n'en voulons pour preuve que ce passage de Albert Kpagodo (1992) : «Les Zémidjan en uniforme à compter du 1^{er} Décembre. Qu'est-ce qui va changer à Cotonou ?»⁴.

Commencé timidement vers 1986, dans les bas quartiers de Cotonou, l'exercice de la profession de Zémidjan⁵ a fait du chemin jusqu'à sa reconnaissance officielle aujourd'hui par les autorités administratives de la circonscription urbaine de Cotonou qui, à défaut de ne pouvoir l'interdire, ont finalement décidé de conjuguer avec, malgré la protestation des propriétaires de taxi et de leurs chauffeurs qui se plaignent de la baisse de leurs recettes journalières... Il est évident que le Zémidjan fait désormais partie de notre vie quotidienne. Mais son organisation est indispensable.

L'État doit promouvoir le développement de langues nationales d'intercommunication. Article 11 de la Constitution de la République du Bénin (décembre 1990).

⁴ *La Gazette du Golfe* est un hebdomadaire indépendant paraissant à Cotonou, n° 55 du 23 Novembre 1992, p. 4.

⁵ Cette orthographe de zémidjan n'est pas conforme à l'alphabet des langues nationales en vigueur au Bénin. Écrire plutôt zemijan. Ici, le tilde est remplacé par (n) pour la nasalisation.

1. DE L'ÉTYMOLOGIE DU TERME ZÉMIDJAN

Selon Nahum (1986), l'origine du terme remonte au début des années 80 à Porto-Novo où une dame n'ayant pas trouvé un taxi, sollicite l'aide d'un motocycliste qui accepte. Mais au moment de monter sur la motocyclette, elle lâche son commentaire comme dans un humour propre aux *wemenu*⁶ :

[zemijanwun die] mot à mot
/prends/ moi/ onomaptopée(qui décrit un geste brutal)/ présentatif (voici)/
traduit : «voici une manière brutale, rigide de me prendre, ou transporter, de monter à califourchon».

Le motocycliste reprend l'énoncé en continuant son chemin. Après cette fortunée, il prit goût à cette activité et en fit une profession.

Mais au départ qui sont-ils ces conducteurs de zemidjans ? Toujours selon Nahum (1986), après le boum pétrolier du Nigéria (deuxième producteur africain de pétrole) des années 70, les autorités nigériennes, sous prétexte que ce sont les immigrants des pays voisins qui font leur malheur, procèdent à un rapatriement massif qu'elles appellent «expulsion des immigrants illégaux».

De retour chez eux avec leur moyen de déplacement, les motocyclettes très à la mode à cette époque et en attendant de trouver du travail, ont enfourché provisoirement leur motocyclette pour pouvoir survivre.

C'est ainsi qu'est née cette activité de conducteur de taxi-moto «zemidjan» qui par la force des choses sera officiellement reconnue comme profession en décembre 1992.

La crise économique des années 80 qui continue de secouer les pays africains, aggravée par la dévaluation du franc CFA a favorisé la multiplication de nouvelles activités professionnelles.

Depuis toujours, ils ne se font pas appeler zémidjanO mais par euphémisme, *kɛkɛn*). Le suffixe *n*) ici est un morphème d'agent ou «marqueur de possession» adjoint à un lexème nominal qui permet de former une nouvelle unité lexicale dans certaines langues déjà étudiées du même continuum. Cette dérivation suffixale est l'un des procédés les plus productifs de création lexicale très usité dans la langue Fon pour la nominalisation.

Pour dénommer les moyens de transport modernes (les automobiles), les fonnu ont emprunté à la langue française le mot taxi [taksi] qui est devenu tasi [tasi] pour être conforme à la structure morphologique du Fongbe qui est une langue à syllabes ouvertes (V, CV). Disons que autrefois, c'est du vocable Gbgamey [gbgamè] que les usagers se servaient pour héler un taxi.

Gbgamey était une région jadis éloignée du centre ville de Cotonou. Alors en plein essor, (installation d'un marché, d'une gare ferroviaire etc.), ce quartier attirait beaucoup de monde. Pour s'y rendre, il fallait un moyen de transport, un véhicule pour

⁶ Population habitant la vallée de l'ouémé, fleuve de la région méridionale du Bénin.

transporter les usagers et leurs bagages. Les premières voitures, pour faire cette course d'un bout à l'autre de la ville, reçurent le nom du quartier où s'est installé le marché.

De Gbgamey, il y a quelques décennies, nous sommes passés à tasi pour désigner tous les véhicules de transport urbains et interurbains impliqués dans cette activité. Ce changement peut s'expliquer par le développement rapide de ces moyens de transport et aussi par le progrès d'une ville moderne comme Cotonou, la porte océane du Bénin.

Le grand marché international Dantokpa de Cotonou est inondé de ces moyens de transport . Les grandes villes du pays comme Cotonou, Porto-Novo (la capitale), Parakou, Ouidah, Abomey connaissent ces moyens de transport, très pratiques pour les courses. Contrairement à l'automobile, le zémidjan vous dépose devant votre portail quel que soit l'endroit où vous habitez et n'importe quand. Vous pouvez négocier le prix de la course à partir de 75 FCFA lorsque la course en ville en taxi commence à partir de 100 FCFA.

2. ZÉMIDJAN : APPROCHE ETHNOLINGUISTIQUE ET SOCIALE

Depuis quelques années, plusieurs tentatives d'organiser les zémidjan pour régler leurs activités ont échoué. Le phénomène zémidjan n'est pas facile à maîtriser parce qu'il recrute la majorité de ses opérateurs parmi les sans-emploi. Dans le département de l'Ouémé, depuis les années 60 et même avant, le vélo jouait le même rôle de vélo-taxi appelé Tasi-kanna. Ce terme est formé à partir de Tasi emprunté à taxi et de kanna pâte de maïs généralement appelé akassa. Ces vélos-homme transportaient la vendeuse d'akassa «Kannanon» et son grand panier. Ces vélos existent encore de nos jours et desservent plus les villages que les villes.

Keke «vélo» est bien connu des populations de toute la partie méridionale du pays. Le terme sans doute est une onomatopée, car d'autres groupes sociolinguistiques comme les yoruba l'utilisent. C'est peut-être à cause du bruit que font les billes lorsque le vélo neuf est en marche ou par imitation phonétique de bicyclette.

Le besoin pressant d'utiliser la langue nationale dans le secteur d'activité nouvellement créé est naturel parce qu'il s'agit de répondre à des nécessités de communication, de formation et de vulgarisation de nouveaux savoir-faire et connaissances.

Pour le cas d'espèce, le public utilisateur de la langue (commerçants, transporteurs, autorités administratives, politiques, médias et étrangers...) se démêle et se perd dans des improvisations, de création spontanée de vocabulaire. En effet, il n'est pas rare d'entendre certains intellectuels les désigner par chauffeur de zémidjan, conducteur de zémidjan ou par les termes comme zémidjantiste, zémidjanman, qui sont des composés hybrides créés à partir du calque morphologique et syntaxique des langues européennes.

Parti d'un énoncé, nous sommes arrivé à l'unité lexicale zémidjan dans toute l'aire culturelle aja-tado où se parlent les langues Gbe. Les populations voisines locutrices des langues Gur comme le Baatonum, ont préféré par apocope de zémidjan, le vocable ze.

Une langue de spécialité se caractérise, rappelons-le avant tout, par la très grande importance des substantifs et de syntagmes nominaux qu'elle comporte. Celui des transports comme le zémidjan en est une illustration.

Nous en avons relevé quelques-uns :

zémidjan	taxi-moto, conducteur de cet engin, personne exerçant cette activité;
kekɛn ʔ	conducteur de motocyclette (terme employé par euphémisme pour désigner un conducteur de zémidjan, motocycliste étymologiquement propriétaire de vélo);
zémidjandot	personne qui emprunte ce moyen de transport;
zémidjann ʔ	propriétaire de zémidjan (le suffixe d'agent nO est un marqueur de responsabilité);
zémidjandote	station de zémidjan;
zokɛke	motocyclette;
zokɛken ʔ	motocycliste;
zémidjanküt ʔ	conducteur de zémidjan.

Comme pour les autres métiers, les nouveaux termes se créent au fur et à mesure que les besoins d'exprimer les pensées se font sentir, et ce en plus des difficultés quotidiennes qui se présentent aux usagers surtout dans les langues à traditions orales comme les nôtres qui ne sont pas encore écrites.

Les utilisateurs de la langue, clientèle potentielle, ont besoin de produits variés, modernisés et implantés dans le vécu langagier de leur communauté linguistique par les canaux appropriés. Cette tâche demande à ce que la langue soit constamment aménagée.

Lors du IV^e Sommet de la francophonie, par exemple, les Cotonnois ont pu lire sur les tee-shirts des conducteurs de taxi-moto, FRANCO ZEM. Depuis décembre 1995, les milieux lettrés en français utilisent le terme zem (qui n'obéit pas à la structure phonologique du Fongbe) pour désigner zémidjan. Notons au passage que cette corporation est exploitée aujourd'hui comme support publicitaire. Les publicistes les utilisent pour lancer un nouveau produit sur le marché. Pendant les campagnes de lutte contre le Sida, les zémidjan sont habillés de la tête au pied et défilent dans toutes les villes et campagnes pour sensibiliser les populations. Ils n'ont pas manqué de participer à leur manière aux différentes élections, législatives, présidentielles, organisées dans le pays.

À propos du terme zémidjan les supports audiovisuels et écrits ont largement contribué à sa vulgarisation. Entre autres actions menées nous citerons :

- comme productions discographiques des albums de disque avec des clips vidéo dont le plus célèbre est intitulé zémidjan de la vedette de la chanson béninoise (Tohon stan 1989).

- la production d'un film documentaire court métrage de 26 minutes du réalisateur béninois Roger Nahum (1986) intitulé : *Zémidjan, taxi-moto de Porto-Novo*.

De nombreux articles dans les journaux de la place continuent d'œuvrer à soigner l'image de marque de ce métier qui, ces derniers temps, est ternie par certaines brebis galeuses.

Le phénomène zémidjan est en pleine expansion.

Partout, sur toute l'étendue du territoire national et même dans les pays limitrophes ce mode de transport fait partie de la vie quotidienne des populations sans distinction d'âge ni de catégorie sociale.

Partout le vocable a pris et a parfaitement intégré les langues des autres groupes socio-culturels du pays tels que le Dendi et le Batonum dans la région septentrionale du Bénin. À Parakou, troisième grande ville du Bénin, les populations préférèrent **Ze**.

Au Togo, pays voisin à l'ouest, plus précisément à Lomé la capitale, ce mode de transport est connu sous le nom de **o le yi a** (énoncé) qui veut dire en Gen-Mina (langue du continuum Gbe) «est-ce que tu t'en vas ?».

Par le biais de cette activité, le transfert de technologie nouvelle en mécanique s'opère par les emprunts aux langues des pays d'origine des pièces de rechange des motocyclettes. Cette activité a favorisé la multiplication des mécaniciens vélo-moteur, de petits forgerons, des «vendeurs de pièces détachées» et vendeurs de motocyclettes deuxièmes main venus du Japon. Elle traduit aussi la dynamique ethnolinguistique des populations Aja-tado, qui de par leurs créativités, inventent à partir d'une technologie importée, un mode de transport adapté à leur milieu sociologique.

Par exemple un lexique spécialisé en mécanique moto est en création et n'est consigné nulle part par écrit dans un document.

Nous présentons ici quelques termes qui sont pour la plupart des emprunts :

buji	bougie
cléapip	clé à pipe
hlévizyɔ	révision
flè avan	frein avant
kekédɔvi	chambre à air

3. DE LA PROBLÉMATIQUE DU MÉTIER DE ZÉMIDJAN

Partout au Bénin, et surtout dans une ville cosmopolite comme Cotonou où la plupart des voies de communication ne sont pas très praticables, le zémidjan passe pour le moyen de transport le plus populaire. Pressés de se faire assez de sous, sans casque de sécurité, sans permis de conduire, ignorant pour la plupart le code de la route, les conducteurs de zémidjan ont pignon sur rue à Cotonou. La réglementation de cette activité constitue aujourd'hui un sérieux casse-tête pour les autorités municipales. Voici quelques chiffres fournis par les services compétents de la municipalité de Cotonou qui parlent d'eux-mêmes.

À Cotonou, ils sont plus de vingt mille. Les points de vente des motos deuxième main «venues du Japon», encore appelées corbillards de par leur couleur noire, se sont multipliés ces dernières années. On en compte plus de 100 dans la seule ville de Cotonou.

Deux types de motos (95 CC et 90 CC) et trois marques japonnaises (Yamaha, Suzuki et Honda) se partagent le marché de ces motocyclettes.

Qui sont aujourd'hui les conducteurs de taxi-moto appelés Zémidjan ?

Bref, tout un monde que ce mode de transport occupe, dépanne et entretient. Parmi eux, beaucoup ne sont pas conducteurs-propriétaires mais travaillent sur contrat d'un an ou un an et demi selon le type et l'état de la motocyclette. Par exemple, nos enquêtes de terrain ont révélé que :

- ce métier emploie surtout des chômeurs, jeunes gens en fin de formation ou en situation d'attente d'un emploi quelconque ou d'un emploi plus rémunérateur. On y trouve aussi les déscolarisés ou de jeunes ruraux en quête de gain facile et même quelques fonctionnaires qui arrondissent leur fin de mois après les heures de service.

Il y a deux modes de gestion de cette activité :

- ou bien deux parties sont liées par un contrat entre un employeur (propriétaire de la motocyclette) et son employé (le conducteur de taxi-moto),
- ou bien le zémidjan est conducteur-propriétaire de taxi-moto.

Avant la dévaluation du franc CFA intervenue, rappelons-le, en janvier 1994, la durée du contrat était de 12 mois. Aujourd'hui elle est de 18 ou 24 mois selon la puissance cylindrique du moteur (50 CC ou 90 CC) avec une recette journalière de 1500FCFA x 6 jours de la semaine. Soit 9 000 FCFA x 4 semaines par mois. Selon le contrat, le conducteur est tenu de verser au total 648 000 FCFA pour le type 50 CC ou 864 000 FCFA pour le type 90 CC selon donc l'état ou la puissance cylindrique de la motocyclette.

Les termes du contrat font obligation au conducteur de la prise en charge de tous les frais (réparation, uniforme, entre autres diverses taxes, le droit taxi est une taxe s'élevant à 600 FCFA par mois). Notons que par jour la recette journalière nette est comprise dans une fourchette de 3 000 FCFA et 4 000 FCFA.

En examinant ces chiffres, nous pouvons dire que le zémidjan est une activité rentable ou tout au moins une épargne qui rapporte 100 % d'intérêt sur l'investissement. Pour le moment, beaucoup de personnes sont attirées par cette activité. On rencontre parfois des propriétaires de deux ou parfois plusieurs taxis-motos en exploitation.

À l'heure actuelle, le phénomène a pris une telle ampleur qu'on ne raisonne plus en termes de le supprimer, mais de l'organiser.

C'est ainsi que depuis 1992, ils sont pris en charge par les municipalités où ils exercent. Ils sont immatriculés à la mairie, portent un uniforme (une blouse jaune affichant leur numéro). Ils sont astreints au paiement de droit de taxi et de vignette comme tous autres véhicules de transport en commun. L'assurance et le permis de conduire (contre lequel ils ont protesté) sont indispensables pour l'exercice de cette activité compte tenu du mauvais état des routes. L'insécurité sur nos routes, dans nos villes et campagnes, et les cas d'accidents graves ont quelquefois porté préjudice à cette profession de conducteur de taxi-moto que certains individus, à tort, pensent relever du secteur informel.

Travailleurs occasionnels, ils se réclament car ils se contentent de cela en attendant d'exercer leur vraie profession. La société les perçoit comme des nécessiteux qui se démêlent pour joindre les deux bouts. L'exercice de cette activité est considéré par les zémidjan eux-mêmes comme transitoire. Bavards, colporteurs des rumeurs de trottoirs, entre deux courses en ville, ils vous racontent presque leur vie, et parfois même vous informent des actualités. Bon nombre d'entre eux prennent l'État pour responsable de leur situation et manifestent à son égard leur ressentiment. Cette attitude peut expliquer l'incivisme notoire qui caractérise certains. La plupart des zémidjan affirment toujours qu'ils y sont pour un temps. Et pourtant, nous en avons rencontrés qui exercent ce métier depuis cinq ans. Parti d'un contrat d'un an, Agossou Codjo⁷, notre informateur de référence, est devenu propriétaire de zémidjan et travaille pour son propre compte. Il entretient sa famille avec un salaire qui tourne autour de 40 000 FCFA par mois tandis que le SMIG⁸ est à 22 000 FCFA. La plupart du temps, ils épargnent entre 500 et 1000F par jour chez des tontiniers⁹. En 1996, le Bénin comptait environ 40 000 zémidjan. Ce chiffre comparé au nombre de fonctionnaires du pays, le dépasse largement. La ville de Cotonou en compte à elle seule plus de la moitié. Nous n'avons pas rencontré des jeunes filles conductrices de taxi-moto. Pourtant, elles aussi sont déflâtées, en chômage et touchées par les maux de la société.

Cela peut s'expliquer par le fait que c'est une activité qui demande beaucoup d'efforts (absence du foyer, endurance à braver toutes sortes de difficultés).

Très tôt le matin, ils sont sur pieds et sont les derniers à se coucher. On rencontre, dans ce corps de métier, toutes les catégories de citoyens. Des ingénieurs, docteurs toutes spécialités confondues qui rentrent de l'étranger après leur formation, exercent cette activité en attendant de s'insérer dans le corps du métier de leur formation.

Beaucoup d'autres emplois se sont créés autour de cette activité. Les essenceries se sont multipliées un peu partout pour leur ravitaillement en carburant et autres produits d'entretien de ces engins à deux roues. Il faut signaler ici que la proximité du Nigéria, d'où rentrent frauduleusement les produits pétroliers et autres pièces de rechange, favorise l'exercice de cette activité. Les zémidjan ne se ravitaillent pas à la pompe où l'essence coûte plus cher.

4. NÉCESSITÉ DE CRÉATION ET DE MISE EN ROUTE D'UN OBSERVATOIRE DES LANGUES NATIONALES AU BÉNIN

Il est urgent pour nos pays de créer un cadre d'aménagement interne des langues. Opérateur officiel des politiques linguistiques, l'Office des langues, parmi ses nombreuses missions et tâches, doit aider à l'installation des offices de langues nationales qui seront des observatoires linguistiques où doit être entrepris le plus tôt possible l'aménagement terminologique au niveau des langues dans nos pays où les programmes d'ajustement structurel favorisent le développement de nouveaux métiers dit «informels». Ce secteur,

⁷ Jeune homme, d'une trentaine d'années, marié et père de deux enfants, électricien ayant exercé ce métier pendant deux ans, originaire du département de l'Atlantique, de coutume Fon, niveau d'étude de la classe de terminale. Actuellement conducteur de taxi-moto.

⁸ Le SMIG est le salaire minimum interprofessionnel garanti.

⁹ Ce sont, dans nos sociétés, des sortes de banques traditionnelles de petits épargnants.

pour s'insérer dans le tissu économique, doit être organisé puisqu'il participe d'une manière ou d'une autre de l'amélioration des situations économiques, sociales et linguistiques de nos populations et partant du pays.

Le marché parallèle de la vente illicite des produits pétroliers n'arrange pas les choses. La lutte qu'a menée l'État contre ce trafic qui constitue un manque à gagner pour l'État s'est soldée par un échec.

Le passage d'une économie planifiée de type socialiste à une économie de marché caractérisée par la libre concurrence et le désengagement de l'État de certains secteurs de production, explique l'essor que connaît cette activité de nos jours.

L'État providentiel du système socialiste dans lequel il ne doit pas y avoir de chômeur (fonction publique pléthorique, création à tour de bras de sociétés d'État) est en train de faire place à une société à économie libérale où les sans-emploi se comptent par milliers.

Bref tout ce monde est recruté par le système D (comme débrouille) qui, en attendant de trouver mieux, font du zémidjan.

Le métier de conducteur de taxi-moto, quoi qu'on dise nourrit son homme au Bénin. Même si l'état des infrastructures routières s'améliorent et que de nouveaux emplois se créent pour diminuer le nombre de chômeur, ce mode de transport, comme son aîné le taxi-kannan, ne disparaîtra point, puisqu'il va continuer à desservir les villages et contrées reculées en voie d'urbanisation. La naissance du syndicat national des conducteurs de taxi-moto est une preuve que l'exercice de cette activité au Bénin ne relève plus désormais du domaine de l'informel, ni de celui des petits métiers de subsistance.

Du 26 au 28 mars 1997, le premier congrès du Syndicat National des Conducteurs de Taxi-moto (SNCTM) s'est tenu à Cotonou où l'on notait la participation d'une délégation des conducteurs de taxi-moto du Togo. À l'issue de leurs travaux, la SNCTM a souhaité son affiliation à la Confédération générale des travailleurs du Bénin. Voilà autant d'indices qui chaque jour ne cessent de renforcer l'idée qu'ils ne doivent plus être considérés comme des marginaux ou des laissés-pour-compte. Les conséquences qui découlent de l'exercice de cette activité sont celles inhérentes à tous corps de métiers ou associations en début de gestation.

Nous assistons déjà à plusieurs formes de déperdition : la déperdition scolaire, celle des jeunes en apprentissage d'un métier, celle des jeunes ruraux pêcheurs et paysans qui abandonnent la terre ou même des étudiants qui désertent parfois les facultés et les écoles pour exercer cette activité. Nous devons signaler les cas de fonctionnaires qui quittent aussi les bureaux quelquefois avant l'heure.

Le rêve de la ville qui attire les jeunes de toutes conditions, et les problèmes de circulation urbaines, la pollution de l'environnement par la fumée, les huiles de vidange, les épaves de véhicules et d'engins de toutes sortes, sont les conséquences de la prolifération des zémidjan.

Vêtus de blouse jaune avec leur immatriculation inscrite en vert au dos (aux couleurs des taxis automobiles au Bénin), les zémidjan s'alignent devant les boutiques, aux abords des marchés, stationnent sur leurs parcs situés non loin des carrefours. Vous les voyez aller quelquefois dans tous les sens faisant parfois fi aussi des règles de la circulation.

À Cotonou, il n'est pas rare de les voir brûler le feu rouge à un carrefour sans agent de police. L'incivisme fiscal se retrouve partout dans tous les corps de métiers et dans toutes les couches de la société et sous tous les cieux. L'État doit s'organiser pour recouvrer ses taxes, mais avant il a le devoir d'assurer l'entretien des infrastructures routières.

Dans l'histoire immédiate de notre pays, le zémidjan traduit la réalité de notre temps avec le manque d'investissement important dans le secteur du transport dans un pays sous programme d'ajustement structurel où l'État providentiel a presque disparu. Aujourd'hui l'État apparaît comme le gendarme qui contrôle tout, beaucoup plus pour asseoir une véritable économie formelle où tous ceux qui échappent au contrôle sont dans l'informel. En réalité, le tissu économique béninois est fait en majorité d'opérateurs de ce secteur. Le concept de secteur informel est un cliché occidental, qu'il faut revoir surtout dans les pays en développement avec leur spécificité sociale et culturelle.

CONCLUSION

Dans la culture des transports au Bénin, la création des nouveaux termes pour désigner les moyens de transports se fait par composition, par affixation par le redoublement ou par emprunt comme le montrent les nouveaux termes créés.

On n'a pas trop de mal à trouver des vocabulaires spécialisés dans nos langues qui utilisent entre autres procédés de création lexicale classiques, la dérivation, la composition par juxtaposition des unités lexicales, les expressions imagées, et surtout le néologisme. Elles sont des langues ouvertes, aptes à exprimer les réalités du monde moderne des sciences et technique. C'est d'ailleurs ce qui a fait dire à Alapini (1950) que le dahoméen (fongbe) est une langue additive.

La connaissance de la technologie traditionnelle africaine à travers les savoirs endogènes : savoir-faire et savoir-être, constitue un préalable indispensable à l'amélioration des savoirs, condition du progrès et partant du développement d'un continent encore à la croisée des chemins. Le développement dont l'Afrique a besoin aujourd'hui doit passer par la prise en compte de son patrimoine culturel en vue de l'adapter aux exigences et besoins d'un monde moderne en perpétuelle mutation.

Ce développement se fera par la langue qui comme l'a si bien dit Stendhal est «le premier instrument par lequel se manifeste le génie d'un peuple.»

La création d'emplois et de projets pour les jeunes en vue de leur auto-emploi peuvent freiner un temps soit peu toutes les formes de déperdition que nous avons observées au cours de notre enquête. Nos langues ne doivent plus être abandonnées à elles-mêmes.

C'est fort de cette vérité que Chumbow (1987) souligne :

«Les langues d'une Nation sont ses ressources économiques (naturelles) au même titre que le pétrole, les minerais et autres ressources naturelles. Elles doivent être exploitées et développées à travers une planification linguistique méticuleuse dans l'intérêt général de la Nation.»

Pour terminer, nous proposons en annexe comme article d'un dictionnaire français le terme zémidjan :

zémidjan [zemijan] n. (1986), origine département de l'Ouémé, du gungbè (langue du Bénin); domaine des transports; motocyclette à places (souvent peint en noir) qu'on peut louer et qui sert de transport et course en ville. Le terme désigne la profession, la motocyclette et quelquefois son conducteur.

RÉFÉRENCES

ACCT et AFELSH, (1997) : *L'année francophone internationale*, AFI, Québec, 348 p.

ALAPINI, J. (1950) : *Le Petit Dahoméen*, Avignon, Les Presse Universelles.

CLAS, A. (dir) (1985) : *Guide de recherche en lexicographie et terminologie*, Paris, ACCT, 155 p.

KPOGODO, M. (1992) : *La Gazette du Golfe*, hebdomadaire indépendant paraissant à Cotonou, n° 55, 23 novembre, p. 4.

NAHUM, R. (1986) : *Les Taxis-motos de Porto-Novo*, Cotonou, ORTB, 26mn (court métrage).

D'HIER À AUJOURD'HUI : ÉVOLUTION ET TENDANCES DE LA NÉOLOGIE MORPHOLOGIQUE DANS LE LANGAGE DE L'INTERNET

Raymond MOPOHO

Université Dalhousie, Halifax, Canada

Le langage de l'ARPANET¹, l'ancêtre de l'Internet, se développe d'abord en anglais et se confond initialement avec celui de l'informatique et de l'électronique. Au fil des ans, il est agrémenté de mots et d'expressions appartenant à ce qu'on peut déjà appeler le jargon informatique, dont les principaux créateurs sont les informaticiens eux-mêmes et, dans une moindre mesure, les écrivains de romans de science-fiction. La dissémination rapide de ce langage — et de la terminologie technoscientifique américaine en général — au niveau international amène plusieurs pays, dont l'Allemagne et la France, à mettre sur pied des organismes de normalisation terminologique. Ainsi, en France le décret n° 72-19 du 7 janvier 1972 relatif à l'enrichissement de la langue française autorise la création par arrêté ministériel de commissions de terminologie chargées :

«d'établir pour un secteur déterminé un inventaire des lacunes du vocabulaire français, [et] de proposer les termes nécessaires soit pour désigner une réalité nouvelle, soit pour remplacer des emprunts indésirables aux langues étrangères» (Journal officiel du 9 janvier 1972).

La création de l'Office de la langue française au Québec s'inscrit au moins en partie dans le cadre d'une démarche similaire. On peut donc constater qu'en anglais le développement du vocabulaire de l'informatique/électronique était essentiellement l'oeuvre des praticiens de ces disciplines, tandis que dans le domaine français il constituait l'activité principale de terminologues officiels. En conséquence, on observait ici une création néologique spontanée, rythmée par la cadence des innovations technologiques, alors que là, on assistait à une création «raisonnée», à la traîne des mêmes innovations, guidée dans une large mesure par la nécessité de lutter contre les emprunts, et recourant abondamment à la traduction et au calque². La vulgarisation des outils informatiques dans les années quatre-vingts, ajoutée à l'avènement et à l'expansion de l'Internet depuis le début des années

¹ «Advanced Research Projects Agency Network», réseau de recherche militaire lancé aux États-Unis en 1969 et dissous en 1990.

² Selon Diki-Kidiri *et al.* (1981 : 12), cette pratique s'explique par le fait que «le terme d'origine étrangère enferme [...] le "néologue" dans une sorte de prison psychologique dont il a la plus grande peine à sortir».

quatre-vingt-dix contribuera à modifier les modes de création néologique dans les deux langues.

En anglais, le jargon des informaticiens, réservé jusque-là à une minorité d'initiés, connaît un enrichissement phénoménal avec l'entrée en scène (vers le milieu des années quatre-vingts) des *hackers*, les mordus de l'informatique. Le langage de l'Internet anglais est largement tributaire de ce jargon dont les néologismes semblent inspirés par une diversité de sources dont la mythologie classique, l'histoire occidentale ancienne et moderne, la culture populaire américaine contemporaine, les sciences exactes et la science-fiction. Toutes ces sources s'entrecroisent sur une toile de fonds où se disputent l'humour, la rébellion contre certains modes de pensée et valeurs traditionnelles, la commisération ou le mépris vis-à-vis de tous ceux pour qui l'informatique ne constitue pas un mode de vie. Ce jargon se caractérise, comme nous l'avons signalé ailleurs (Mopoho 1996 : 130), par le doublage des verbes, l'application de la dérivation nominale à tous les verbes et de la dérivation verbale à tous les noms, l'agglutination de termes phonologiquement apparentés, le rejet des règles grammaticales et la généralisation de leurs exceptions³, le recours à des complexes consonantiques difficilement prononçables, l'utilisation abondante de la métaphore, l'application de la siglaison⁴ aux expressions fréquentes de la langue courante (**A+** pour *à plus tard*, **MDR** pour *mort de rire* qui est l'équivalent du sigle anglais **ROFL**, *Rolling on the floor laughing*), etc.

En français, la vulgarisation de l'Internet a été suivie par une certaine période de flottement au cours de laquelle les terminologues officiels paraissaient dépassés par les événements, en raison de la grande rapidité des innovations. Pendant ce temps, l'engouement suscité par le nouveau médium contribuait à grossir les rangs des usagers plus ou moins initiés en informatique qui n'hésitaient pas à recourir à des emprunts massifs pour résoudre leurs nombreux problèmes de vocabulaire. Après cette période initiale, on a assisté à l'apparition d'une classe de mordus francophones qui, comme leurs homologues anglophones (les *hackers*), allient la création néologique à la pratique professionnelle. Certains d'entre eux vont, individuellement ou collectivement, créer des bases de données unilingues ou multilingues qui sont mises à la disposition (sans frais) du grand public dans des sites Web. Quelques organismes officiels de terminologie et de normalisation réagissent en proposant eux aussi sur l'Internet des bases de données du même genre. Ces glossaires, lexiques et dictionnaires, dont la qualité trahit l'amateurisme ou reflète le professionnalisme des divers auteurs, ont tous un point en commun : le caractère innovateur des procédés de création néologique. S'il est évident que dans le domaine français on fait appel aux modes de formation aussi bien agglutinogène que syntaxique, le principe d'économie semble jouer un rôle prépondérant partout, à cause probablement de la brièveté des concurrents que sont les emprunts anglais. En outre, la composition apparaît comme l'un des procédés les plus utilisés dans le mode de formation syntaxique. L'hybridation, fort répandue dans la formation néologique en informatique/électronique, ne s'arrête plus à la combinaison de formants grecs et latins, mais aussi à celle de l'anglais et de l'allemand (*infobahn* par exemple). Mais l'innovation

³ Cette pratique est résumée et illustrée par la formule «All verbs must be nounced and all nouns must be verbed».

⁴ Cet engouement pour les sigles peut s'attribuer à la fois aux origines militaires de l'Internet (cf. note 1 ci-dessus) et à la recherche de la concision chez les informaticiens.

qui est à notre avis la plus importante et qui se trouve au centre du présent article est la suivante, l'apparition de nouveaux formants.

LES NOUVEAUX FORMANTS

Le dépouillement d'un corpus restreint d'environ un millier d'expressions et mots fréquemment employés par les participants aux forums électroniques ou figurant dans les lexiques disponibles sur l'Internet nous a permis d'identifier un certain nombre de formants qui rentrent dans la création de la plupart des néologismes. Pour établir une typologie sommaire de ces formants, nous avons songé dans un premier temps à distinguer entre les formants monématiques et les formants complexes. Mais nous avons dû abandonner cette classification parce que, entre autres, certains formants présentent des allomorphes comme on le verra plus loin. Nous avons également pensé à les répartir en formants dérivatifs et en lexies (Diki-Kidiri *et al.* 1981), mais nous avons également écarté cette distinction parce que les dérivatifs ne présentent aucun caractère de nouveauté. À l'inverse, il s'avère que la distinction entre confixes et pseudo-confixes s'accommode mieux avec l'objectif de notre analyse.

À titre de rappel, les confixes sont des racines liées d'origine grecque ou latine, tandis que les pseudo-confixes sont des mots tronqués qui rentrent dans la formation de termes composés (Kocourek 1991 : 130). Dans le cadre de la présente description, nous avons expressément inclus dans la classe des pseudo-confixes les formants qui résultent de la simplification d'unités syntagmatiques ou de la siglaison. Nous avons ainsi répertorié une trentaine de pseudo-confixes qui nous semblent nouveaux et qui rentrent dans la composition de nombreux néologismes, et nous les avons classés de la manière suivante :

a) Les pseudo-confixes antérieurs

Il s'agit de formants qui résultent généralement de l'apocope de certains termes et dont la liste est la suivante :

- e-** < électronique (*e-zine*)
- é-** < électronique (*étexte*)
- É-** < électronique (*É-zine*)
- cyb-** < cybernétique (*cybionte; cyborg*)
- cyber-** < cybernétique (*cybercafé; cyberpolice*)
- info-** < information/informatique (*inforoute; infogérance*)
- M-** < méga- (*Mbit; Mbps*)
- n-** < network (*néthique; nétiquette*)
- net-** < internet (*netsurfeur; netPol*)
- nul-** < nul(le) (*nultilisateur; nultilisatrice*)
- P-** < pseudo- (*P-code*)
- publi-** < publicité (*publipostage*)
- web-** < World Wide Web (*webmestre; webnaute*)

b) Les pseudo-confixes postérieurs

Contrairement aux pseudo-confixes précédents, ceux-ci résultent en général de l'aphérèse, voire de l'élagage des composants d'une unité syntagmatique, à l'exception du dernier élément :

- el < électronique (*courriel; babel*)
- iciel < logiciel (*gratuiiciel; partagiciel*)
- icien < informaticien (*qualiticien; imprimaticien*)
- iel < logiciel (*contributiel; tutoriel*)
- ique < informatique/électronique (*bancatique; éditique*)
- naute < astronaute (*internaute; infonaute*)
- net < Internet (*intranet; supernet*)
- zine < magazine (*webzine; e-zine*)

c) Les pseudo-confixes sémantiques

Ces formants sont des homographes de confixes classiques, mais ils ont un sens différent de celui de ces derniers, soit à cause d'un glissement sémantique, soit parce qu'ils résultent de la troncation (surtout par apocope) de termes autonomes :

- anté-** < anté- (*antémémoire*)
- anti-** < anti- (*antiprogramme*)
- hyper-** <hypertexte (*hyperdocument; hyperimage*)
- inter-** <internet (*internouille; interpédia*)
- macro-** < macrocommande (*macro-instruction*)
- micro-** < micro-ordinateur ou micro-informatique (*microédition; micromiser*)
- nano-** < nano- (*nanotechnologie; nanomachine*)
- rétro-** < rétroaction ou rétrograde (*rétro-modération; rétro(-)technique*)
- super-** < super- (*superutilisateur*)
- téra-** < téra- (*téraflop*)
- trans-** < transistor (*transputer; translateur*)

CARACTÉRISTIQUES DES PSEUDO-CONFIXES

La caractérisation rigoureuse des pseudo-confixes susévoqués nécessite une description morphologique et sémantique exhaustive qui reste malheureusement à entreprendre. En attendant, nous nous contenterons pour le moment de faire des remarques générales inspirées par une analyse succincte de quelques-uns de ces nouveaux formants.

Des pseudo-confixes qui ont la même forme et en général la même signification en anglais et en français peuvent provenir de sources directes différentes. Le formant français **cyber-** vient de *cybernétique* et signifie «propre au monde virtuel de l'Internet», tandis que son homographe et équivalent anglais vient de *cyberspace*⁵, néologisme inspiré apparemment de *cybernetics*. En outre, en français **cyber-** (qui est synonyme du pseudo-confixe **info-** dans certains contextes : *cybernaute = infonaute*) apparaît tantôt comme un morphème lié, et tantôt comme une unité autonome. Ainsi, on trouve à la fois *culture cyber* et *cyberculture*.

⁵ On fait généralement remonter la première attestation de ce terme à 1984. Cependant, dans le *New York Time Magazine* en date du 14 juillet 1996, son créateur William Gibson affirme l'avoir employé pour la première fois dans un article paru en 1981, avant de l'utiliser dans son roman de science-fiction *Neuromancer* publié en 1984.

Alors que dans l'exemple de **cyber-** le même formant provient de deux sources directes distinctes en français et en anglais, deux mots français différents peuvent être à l'origine d'un même formant. C'est le cas du pseudo - confixe **info-**, qui résulte de la troncation aussi bien de *information* que de *informatique*. Il en découle par exemple que *inforoute* signifie «autoroute de l'information», alors que *infogérance* renvoie à la «gestion en sous-traitance d'un système informatique». Seul le contexte permet de déterminer la nature de la source abrégée.

Certains pseudo-confixes dont l'existence remonte à la période pré-Internet voient s'accroître leur productivité. Nous pensons particulièrement aux formants **-ique** (aphérèse de «électronique» ou «informatique») et **-iciel** (aphérèse de «logiciel»). Dans le cas de **-ique**, l'augmentation de la productivité est due à l'extension sémantique : au sens traditionnel de «application de l'électronique et l'informatique à» s'ajoutent désormais ceux de «technique de», «propre à», et «ensemble des techniques appliquées à», comme l'illustrent respectivement les néologismes *animatique*, *escargotique* et *connectique*. En ce qui concerne le formant **-iciel**, sa popularité peut s'expliquer par l'intensification des activités de programmation et la multiplication de ses allomorphes, comme le montrent les syllabes finales de *contributiel* (synonyme de *partagiciel*), *progiciel* (logiciel professionnel) et *tutoriel* (synonyme de *didacticiel*).

La plupart des pseudo-confixes sémantiques sont utilisés à la fois dans leur sens traditionnel et dans leur nouvelle acception. On a par exemple des usages parallèles dans **hyper-** («très au-dessus (de) vs «interactif»), **inter-** («entre» vs «de l'Internet»), **macro-** («grand» vs «macrocommande»), **micro-** («petit» vs «application de la micro-informatique»), et **super-** («au-dessus (de)» vs «privilegié ou exceptionnel»).

Le pseudo-confixe obtenu par apocope occupe une position préfixale, tandis que celui résultant de l'aphérèse se trouve dans une position suffixale. Cependant, dans la structure du composé, la fonction du formant postérieur est indépendante de sa position. Les formants **-e**, **-el** et **-ique** sont généralement des déterminants, tandis que **-iciel**, **-naute** et **-zine** sont des déterminés. Par contre, hormis **M-**, les formants antérieurs sont des déterminants, tout comme les pseudo-confixes sémantiques. Le formant **(-)net(-)** constitue un cas d'exception puisqu'il peut être déterminant ou déterminé, selon sa place dans le composé. En général, les formants antérieurs et postérieurs conservent le sens des termes dont ils sont l'abrègement.

En ce qui concerne la soudure des éléments de composition, elle se fait systématiquement sans ajout d'une voyelle de liaison. Cette pratique peut s'expliquer par le fait que l'introduction de la voyelle de liaison répond à des besoins d'euphonie. Or, cette dernière est de peu de pertinence dans le discours de l'Internet où le texte est surtout destiné à être parcouru des yeux et non articulé. À cette raison, on peut ajouter la culture prédominante des *hackers* qui, comme nous l'avons déjà mentionné, privilégient nettement la création de formes imprononçables. En plus, en raison de l'instabilité de la morphologie des nouveaux formants, les points de coupure tendent à changer en fonction des mots de base : **-iciel** se contracte en **-iel** dans *tutoriel* et s'étend en **-giciel** dans *progiciel*, tandis que *Internet* se décompose en cinq pseudo-confixes correspondant à autant d'allomorphes : **inter-** (*internouille*), **intern-** (*internite*), **-net** (*supernet*), **net-** (*netsurfeur*) et **n-** (*nétiquette*). Par ailleurs, le mot de base qui reçoit le pseudo-confixe peut lui-même être tronqué, comme dans *Netpol* ou *e-zine* où les mots *police* et *magazine* ont été abrégés.

SYNONYMIE ET ÉQUIVALENCE

Bien que dans la création de la plupart des néologies on recourt de plus en plus à l'adaptation, dans l'ensemble le procédé de formation prédominant reste le calque de structure, de sens et de métaphore, où les nouveaux formants jouent un rôle clé. Sur la base des néologismes et de leurs signifiés courants, on peut établir (s'il y a lieu) entre les formants susévoqués des relations de synonymie et d'équivalence (par rapport à l'anglais) comme suit :

Pseudo-confixe	Équivalent anglais
anté- Syn. cache caché -e	cache Syn. cached
anti- Syn. inverse opposé -e	back- Syn. reverse
cyb- (voir cyber-)	
cyber- Syn. branché -e cyber électronique info- interactif -ve Internet -net- numérique télé- virtuel, -le	cyber- Syn. cyberspace digi(t) digital electronic Internet n- net- remote virtual wired
e- Syn. é- É- automatisé -e automatique el- électronique robot	e- Syn. electronic automated automatic -bot robot
é- (voir e-)	
É- (voir e-)	
-el Syn. électronique	e- Syn. electronic wired automated

*D'hier à aujourd'hui : évolution et tendances de la néologie morphologique
dans le langage de l'internet*

hyper- Syn. hypertexte interactif	hyper- Syn. clickable
-iciel Syn. iel logiciel	-ware
-icien Syn. informaticien, -ienne futuriste de l'an 2000	
-iel (voir -iciel)	
info- Syn. branché -e électronique information informatique automatique automatisé -e	info- Syn. cyber electronic I- information news- wired
inter- Syn. info- Internet net-	inter- Syn. Internet -net
intern- (voir inter-)	
-ique Syn. électronique informatique technique numérique -tique	e- Syn. digi(t)- digital electronic
M- Syn. méga	M- Syn. mega
macro- Syn. macrocommande	macro- Syn. macrocommand
micro- Syn. microordinateur électronique informatique	micro- Syn. e- electronic

-n Syn. cyber- net- Internet virtuel(le)	n- Syn. Internet net- virtual
nano-	nano-
-naute Syn. -routard -routier	-naut Syn. -izen
-net (voir net-)	-net
net- Syn. cyber- Internet -net	net- Syn. cyber- Internet
nul- Syn. amateur non-initié -e	loser Syn. l-
P- Syn. pseudo illusoire artificiel -le	P- Syn. pseudo fake artificial
publi- Syn. publicité publicitaire	-mercial Syn. commercial ad
rétro- Syn. inverse contraire	retro- Syn. ret- reverse
super Syn. exceptionnel -le	super Syn. exceptional
téra-	tera-
trans-	trans-
Web- Syn. W3 W cube Web WWW	web- Syn. World Wide Web www

<p>-zine Syn. magazine publication journal</p>	<p>-zine Syn. magazine publication journal</p>
---	---

PRODUCTIVITÉ ET VIABILITÉ

Pour le moment, tous les nouveaux formants identifiés ci-dessus n'ont pas la même fréquence d'apparition dans les néologismes de l'Internet. Certains, tels que **M-**, **macro-** ou **anté-** entrent dans la composition de quelques termes seulement⁶. D'autres par contre, à l'instar de **cyber-**, **-naute**, **-ique**, **info-**, **web-**, **inter-**, **-iciel** et **hyper-** reviennent fréquemment dans la plupart des nouveaux composés et ils relèguent au second plan tous les autres formants, y compris confixes classiques dont les plus usités sont **audio-**, **auto-**, **bi-**, **co-**, **de-/dé(s)-**, **intra-**, **méga-**⁷, **méta-**, **multi-**, **pré-**, **re-/ré-**, **sous-**, **sub-**, **sur-**, **vidéo**. Ainsi que nous l'avons eu l'occasion de le relever (Mopoho 1997 : 123-124), la productivité (au sens de possibilité de dérivation) de chaque nouveau terme dépend principalement de sa syntaxe. Dans notre corpus, peu de composés dont le déterminant vient en position finale font l'objet d'une dérivation, celle-ci étant généralement du type verbal, comme dans l'exemple de *courrieler* (*courriel*). Par contre, nous avons constaté que quand le déterminant est placé en position initiale dans le composé, ce dernier se prête à de nombreuses dérivations : *inforoute* permet ou pourrait permettre la formation de *inforoutier/inforoutière* (substantif ou adjectif qualificatif), *inforoutard* (substantif), *inforouter* (verbe) et *inforoutièrement* (adverbe). La structure **déterminé + déterminant** engendre des néologismes d'une grande maniabilité. Le fait qu'une telle structure enfreigne les règles de la composition savante en français (cf. Zwanenberg 1987) serait dû à des raisons pratiques, car la plupart des syntagmes lexicalisés de l'Internet (*logiciel à coût partagé*, *virgule flottante*, *moteur de recherche*, *site miroir*, *robot d'annulation*, *nom de domaine*, *numéro local d'accès*, etc.) sont conformes à la syntaxe de la langue française.

Qu'en est-il de la viabilité des termes formés avec les nouveaux pseudo-confixes ? Il est sans doute trop tôt pour apporter une réponse satisfaisante et définitive à cette question. L'observation suggère cependant que l'implantation des néologismes dans le discours de l'Internet est relativement rapide : les nouveaux termes semblent entrer dans l'usage au fur et à mesure qu'ils sont formés. La grande vitesse de cette implantation reflète, encore une fois, la nature intrinsèque de l'univers de l'Internet où tout se déroule très vite.

⁶ En cela ils diffèrent peu des pseudo-confixes ad hoc qu'on ne trouve que dans un seul terme, comme **liber-** de *libertel* (équivalent de *freenet*), **pro-** de *progiciel* (logiciel professionnel) ou **PC-** de *PC-cillin* (synonyme de *antivirus*). Ce genre pseudo-confixes est plus fréquent dans le domaine anglais : **M-** < *mail* > *Mbox*, **E-** < *European* > *Ebackbone*, **S-** < *snail* > *Smail*, **key-** < *keyboard* > *keypal*, etc.

⁷ Celui-ci est utilisé à la fois comme préfixe dans les noms d'unités physiques (pour indiquer une valeur de 10⁶) et comme préfixe intensif pour signifier «de très grande dimension», par opposition à **téra-** qui veut dire «de taille incommensurable, inouïe».

CONCLUSION

Il serait évidemment erroné d'affirmer que le phénomène décrit ci-dessus, à savoir l'utilisation de pseudo-confixes à des fins de création néologique, constitue en elle-même une nouveauté. Dans l'industrie du divertissement et de la publicité commerciale, les entreprises ont depuis longtemps pris l'habitude de créer et, pour ainsi dire, de se disputer des «bouts de mot» qui servent à désigner des noms de marques et des activités. On remarquera ainsi, par exemple, que dans l'audio-visuel et les sports, le pseudo-confixe **-thon** (de *marathon*) a le sens de «événement de durée exceptionnellement longue, organisée à des fins caritatives», d'où les composés *téléthon*, *radiothon*, *marchothon*, etc. De même, le formant **-tel** s'avère très prisé aussi bien dans l'industrie hôtelière (le mot tronqué étant «hôtel» : Sofitel, Novotel) que dans celle des télécommunications (Sasktel, BCTel, Nortel, Cantel...). La nouveauté réside spécifiquement dans l'ampleur du phénomène en question, caractérisée par l'apparition d'un nombre relativement élevé de pseudo-confixes, la richesse et la diversité des nouveaux composés, la grande vitesse de vulgarisation de ces derniers, et le type de discours (plus général) dans lesquels ils sont utilisés. En ce qui concerne la «philosophie» qui sous-tend les activités de formation de néologismes, on observe une certaine différence entre le français et l'anglais : contrairement aux *hackers* chez qui la création néologique est surtout guidée par la volonté de bâtir un jargon incompréhensible pour les non-initiés, les mordus du domaine français semblent davantage portés à créer des termes accessibles au grand public. Cette volonté est toutefois compromise par la tendance encore répandue qui consiste à reproduire en français les procédés néologiques du domaine anglais.

RÉFÉRENCES

Commission ministérielle de terminologie informatique (France) : *Glossaire des termes informatiques*,

[http://www.rocq.inria.fr/_deschamps/www/CMTI/glossaire.html]

DIKI-KIDIRI, M., JOLY, H. et C. MURCIA (1981) : *Guide de la néologie*, Paris, Conseil international de la langue française.

Glossaire de l'ORSTOM [http://www.pasteur.fr/other/computer/unix/unixguide_html]

Glossaire des termes relatifs à l'Internet [<http://www.culturefr/dglf/internet.htm>]

Le glossaire [<http://www.neomedia.com/iabdd/glossair/glossair.htm>]

Le Jargon Français [<http://www.mygale.org/05/jargonf>]

KOCOUREK, Rostislav (1991) : *La langue française de la technique et de la science*, deuxième édition revue et augmentée, Wiesbaden, Oscar Bandstetter Verlag GMBH & CO. KG.

Lexique bilingue de l'Internet [<http://www.tpsgc.gc.ca/termium/text/vocab-f.html>]

Lexique de la cyberculture [<http://www.queim.fr/CSphere/Lexique.html>]

Lexique des néologismes Internet

[<http://www.ina.fr/cp/HumainsAssociés/Lexique/Lexique.html>]

MOPOHO, Raymond (1996) : «L'emprunt et la créativité lexicale dans le discours de l'Internet français», *Alfa*, 9, pp. 123-132.

MOPOHO, Raymond (1997) : «Composition et traduction dans la terminologie de l'Internet», *Mélanges de linguistique offerts à Rostislav Kocourek*, Halifax, Les Presses d'Alfa, pp. 119-124.

NETGLOS [<http://www.li.com/Translation/Netglos/Glossary>]

Office de la langue française (Québec) : *Vocabulaire d'Internet*

[http://www.olf.gouv.qc.ca/doc_glos.html]

Vocabulaire de l'inforoute [<http://w3.cidif.org/naviguer/vocabulaire.htm>]

ZWANENBERG, Wiecher (1987) : «Le statut de la formation des mots savants en français et en anglais», *Meta*, 32-3, pp. 223-229.

QUELQUES MUANCES ET UNE DÉFINITION DU MOT «ACCENT»

Charles DOUTRELEPONT

Université Carleton, Ottawa, Canada

Les définitions, l'analyse et le classement des sens pour un terme donné varient selon que l'on consulte tel ou tel dictionnaire (Quémada 1967 : 389-500); il en va ainsi des définitions de l'«accent». Par exemple, sous cette rubrique, le *Dictionnaire alphabétique et analogique de la langue française* (Rey 1989 : I, 53) recense et définit un certain nombre de phénomènes relatifs à l'accent. En fonction d'affinités sémantiques, les définitions qui y sont proposées sont réparties en trois groupes (correspondant aux chiffres romains I, II, III). Le premier groupe rassemble les définitions suivantes de l'accent :

- 1 - «Augmentation d'intensité de la voix sur un son, dans la parole [...]»;
- 2 - *a*) «Élévation de la voix sur un son (accent de hauteur; accent musical)», *b*) «Accent de phrase» [...];
- 3 - «Signe graphique qui note un accent [...] (En français) Signe qui, placé sur une voyelle, la définit [...]».

Le deuxième groupe comprend cinq autres définitions :

- 1 - «Ensemble des inflexions de la voix (timbre, intensité) permettant d'exprimer les sentiments, les émotions [...]»;
- 2 - «Fig. Caractère personnel (du style), manière caractéristique [...]»;
- 3 - «Par anal. *Un accent* : intensité plus forte d'une couleur, d'un trait [...]»;
- 4 - «Loc. Fig. *Mettre l'accent sur quelque chose*» [...];
- 5 - «Littér. *Les accents* : sons expressifs» [...].

Le troisième groupe correspond à une seule définition de l'accent :

- «Ensemble des caractères phonétiques distinctifs d'une communauté linguistique considérés comme un écart par rapport à la norme, dans une langue donnée [...]».

Toujours sous la rubrique de l'accent, alors que, comme on vient de le voir, le *Dictionnaire alphabétique et analogique de la langue française* distingue trois catégories de phénomènes, le *Grand Dictionnaire Larousse* en recense deux (Larousse 1987 : I, 25). La première catégorie correspond aux définitions suivantes :

- 1 - «Élévation de la voix portant sur une des syllabes d'un mot ou d'un groupe de mots [...]»;
- 2 - «Intensité particulière avec laquelle un chanteur ou un musicien exécute certaines notes de la phrase musicale [...]»;

- 3 - «Façon de prononcer une langue, inflexions de la voix propres aux hommes d'un pays, d'une province [...]»;
- 4 - «Insistance avec laquelle est indiqué le trait dans un dessin [...]»;
- 5 - «*Mettre l'accent sur quelque chose*, y insister [...]»;
- 6 - «Intonation de la voix exprimant un sentiment, une pensée [...]»;
- 7 - «Marque caractéristique [...]».

La deuxième catégorie est définie en ces termes :

- «Signe graphique placé au-dessus d'une voyelle pour en préciser la valeur».

Si on compare donc les articles que le *Dictionnaire alphabétique et analogique de la langue française* et le *Grand Dictionnaire Larousse* consacrent à l'accent, on se rend compte à la simple lecture que les définitions de l'accent ne concordent pas. L'examen d'autres répertoires, anciens et contemporains, dictionnaires ou encyclopédies, confirmerait les variations dans les perspectives. À côté de ces divergences, dans les définitions et dans les procédés définitoires, qui organisent donc le savoir différemment d'un répertoire à l'autre, s'observe cependant une convergence : dans bon nombre de répertoires anciens et contemporains, l'«accent» est défini comme une «élévation de la voix». C'est cette définition qui m'intéresse; mon hypothèse de travail est qu'elle ne date pas d'hier.

Je me propose d'abord de montrer l'importance de cette définition de l'accent, en la traquant dans les répertoires du français, contemporains et anciens, et en examinant quelques-unes de ses nuances. J'examinerai ensuite en diachronie, sans prétention d'exhaustivité, un certain nombre des définitions de l'*accentus*, en latin classique d'abord, puis en latin médiéval. Pour l'accent, des fragments des définitions latines et grecques ont passé dans les premiers répertoires du français. Revoir ces vieux textes, c'est donc éclairer les définitions que retiennent les premiers lexicographes du français. Ces définitions de l'accent — et c'est là sans doute que se situe la dimension la plus vivante de la mémoire du mot — informent encore aujourd'hui, à des degrés divers, les définitions des dictionnaires contemporains.

Avant d'aborder les définitions de l'«accent» dans les dictionnaires et les encyclopédies, je rappellerai que depuis une trentaine d'années, les linguistes définissent l'accent comme une «mise en relief¹», d'une unité accentuable — en français, la syllabe —, dans le cadre d'une unité accentuelle — en français, le mot ou le groupe de mots (Garde 1968 : 27-31). L'accent se présente comme :

«une proéminence d'énergie articulatoire, qui se manifesterait par une augmentation physique de *longueur*, d'*intensité* et éventuellement un changement de *fréquences* en passant de syllabe inaccentuée à accentuée et/ou au cours de l'évolution de la syllabe accentuée» (Léon 1992 : 107).

C'est sous cet angle que l'accent est défini dans l'*Encyclopædia Universalis* (1990, A-Dieu : 25) :

* Je remercie Isabelle Leroy-Turcan et Xavier Lelubre pour leurs suggestions. Le texte n'engage que moi.

¹ Voir à ce sujet, entre autres, les travaux de Martinet (1968); Garde (1968 : 31), Carton (1972: 98-103), Allen (1973 : 86-95), Clas (1982 : 113-117), Léon (1992 : 1-7-109).

«en général, il s'agit de phénomènes d'ordre phonique qui consistent à donner un certain relief à telle ou telle partie de la chaîne parlée (syllabe, mot, syntagme, etc.) en faisant contraster un segment particulier avec le reste de l'énoncé. Cette mise en relief peut être réalisée par des moyens variés qui se présentent toujours comme plus ou moins liés entre eux : intensité, durée, hauteur.»

Or, ce point de vue — qui se situe dans la perspective des travaux des linguistes des trente dernières années — se distingue de celui de plusieurs dictionnaires contemporains.

En effet, dans plusieurs² répertoires du XX^e siècle, l'accent est défini comme une «élévation de la voix». Cette définition, dont la dimension métaphorique est évidente, en ce que la voix ne s'élève pas, est une des définitions du *Dictionnaire alphabétique et analogique de la langue française* (Rey 1989 : I, 53); c'est aussi la première définition du *Grand Dictionnaire Larousse* (1987 : I, 25). Cette définition de l'accent comme une «élévation de la voix», elle se présente aussi, en tout début d'article, dans les répertoires suivants :

- le *Dictionnaire alphabétique et analogique de la langue française* (Robert 1971 : I, 24; 1957, I, 24);
- le *Dictionnaire étymologique et historique du français* (Dauzat et alii 1964 : 5);
- le *Dictionnaire de l'Académie française* (Académie 1932 : 10);
- le *Dictionnaire général de la langue française* (Hatzfeld et Darmesteter 1924 : 2);
- le *Dictionnaire de la langue française* (Littré 1885 : I, 27; *idem* 1873 : I, 27).

Si on remonte le cours du temps, au XIX^e siècle, on constate que l'accent n'a pas été défini uniquement comme une «élévation» de la voix, mais aussi comme un «abaissement» de la voix; à la manière d'un habitant des montagnes, la voix pourrait monter ou descendre. L'accent est défini comme «élévation ou abaissement de la voix» par les dictionnaires suivants³ :

- le *Nouveau Larousse illustré* (Augé 1898 : I, 6);
- le *Dictionnaire de l'Académie française* (Académie 1878 : I, 14; *idem* 1835 : I, 10; *idem* 1798);
- le *Grand Dictionnaire universel du XIX^e siècle* (Larousse 1866 : I, 7).

Au XVIII^e siècle, c'est par des formulations approchantes que l'accent est encore défini. Je citerai le début de la définition du *Dictionnaire de l'Académie française* (Académie 1762 : I, 25) : «Élévation plus ou moins forte de la voix sur certaines syllabes [...]». C'est encore par une formulation voisine que débute l'article que Richelet dans son *Dictionnaire de la langue française ancienne et moderne* consacre à l'«accent»; comme cet article de Richelet m'apparaît essentiel, j'en citerai une large part en modernisant quelque peu l'orthographe (1681 : I, 22) :

² J'évite de m'avancer plus avant n'ayant pas pu examiner systématiquement tous les répertoires. Il faut signaler aussi que le terme «accent» ne figure pas dans plusieurs répertoires anciens, entre autres dans les *Curiositez françoises* de Oudin (1640), dans le *Dictionnaire étymologique ou Origines de la langue françoise* de Ménage (1644).

³ Je relève encore une définition voisine, celle du *Nouveau Dictionnaire national* (Bescherelle 1893 : I, 37; *idem* 1874 : I, 39; *idem* 1867 : 38) : «Modulation de la voix humaine qui s'élève ou qui s'abaisse sur certaines syllabes».

«Accent. Terme de grammaire que les Grecs et les latins appellent *tóyow*, ton, parce qu'il sert à relever ou à rabaisser la voix en prononçant un mot. Il y a trois sortes d'accents : l'aigu, le grave et le circonflexe. Les deux premiers sont simples, et le troisième est composé des deux autres. L'aigu est une petite ligne tirée de la droite à la gauche, (´), le grave, au contraire, est une ligne tirée de la gauche à la droite, (̀) le circonflexe est composé des deux autres accents (^).

L'aigu marque qu'il faut élever la voix; le grave qu'il faut l'abaisser; et le circonflexe qu'il faut tenir un milieu.»

L'auteur de ce dictionnaire postule aussi que l'accent correspond à deux modalités vocales : élever ou relever la voix, et abaisser ou rabaisser la voix. Richelet dans la description des trois accents du français recense cependant une troisième modalité : la voix pourrait «tenir un milieu». Richelet indique en outre comment dessiner les accents graphiques. Enfin, pour Richelet, il y a une correspondance entre la voix et les signes auxiliaires de l'orthographe : «l'aigu marque qu'il faut élever la voix; le grave qu'il faut l'abaisser; et le circonflexe qu'il faut tenir un milieu». Cette correspondance, j'y consacre les pages suivantes, est ancienne.

En remontant au plus près des commencements de la lexicographie française, au *Thresor de la langue françoise* de Nicot (1621 : 6), on constate que, pour Nicot, l'accent est susceptible de correspondre à trois modalités de la voix : montée, descente et — combinaison des deux premières modalités — montée et descente. Nicot écrit à propos de l'accent⁴ :

«Est pur latin et signifie l'élévation, ou rabaissement, ou contour de la voix en prononçant quelque diction, *Accentus*, et conséquemment signifie les virgules et marques apposées aux mots indiquant les endroits d'iceux où il faut hausser, ou rabaisser, ou contourner la voix : dont il y a trois manières, accent aigu, dont voici la figure, ´, accent grief ou grave, ` , et accent circonflexe ou contourné, ^ ou ~.»

Pour Nicot, aux trois modalités de la voix, élévation, rabaissement et contour, correspondent aussi trois accents graphiques : l'aigu, le grave ou grief, et le circonflexe. L'accent aigu indique que la voix s'élève, l'accent grave qu'elle s'abaisse, le circonflexe qu'elle est contournée.

En fonction de la tradition décrite jusqu'ici, cet examen sommaire, en diachronie, permet déjà de préciser quelques-unes des nuances de l'accent. Ainsi, entre le XVII^e siècle et le XX^e siècle, le rapport entre les accents graphiques et les modalités de la voix disparaît; la définition de l'accent comme «élévation de la voix», qui caractérisait l'accent aigu, semble par contre avoir été maintenue pour définir globalement l'accent. Or, quand Nicot, au XVII^e siècle, définit l'accent, passent aussi en français des fragments des définitions que les grammairiens latins, à partir surtout du IV^e siècle de notre ère, avaient élaborés pour le terme latin correspondant *accentus*. Il en est ainsi en particulier de la correspondance entre signes diacritiques et modalités de la voix.

⁴ Je modernise la graphie et la ponctuation.

Pour situer l'émergence du mot latin *accentus*, et ses définitions, dans les grammaires latines du IV^e et du V^e siècles de notre ère, je présenterai rapidement les spéculations théoriques de l'antiquité qui concernent la nature du son, celles de Platon en particulier. Je rappellerai ensuite l'élaboration par les savants de l'école alexandrine d'un système de ponctuation. Il faut bien dire que, pendant longtemps, à travers l'histoire, le terme latin *accentus* et, par la suite, le terme français «accent» ont été définis à la fois en fonction de spéculations abstraites sur les sons et de considérations concernant les signes auxiliaires de l'orthographe.

Parmi les premières spéculations sur la nature des sons qu'il y a lieu d'évoquer, il faut signaler le *Cratyle* de Platon, au IV^e siècle avant notre ère. Dans ce dialogue, Platon attribue aux sons une valeur expressive : certains sons suggèrent le «mouvement», d'autres l'«arrêt» (*Cratyle* : 426 d - 427 c; cf. Collart 1968 : 73-75). Les sons, par ailleurs, y sont classés selon une triple nomenclature : voyelles, semi-voyelles ou intermédiaires, et consonnes muettes (Collart 1968 : 75). Enfin, et c'est ce qui m'importe, Platon (*Cratyle*, 399 a) distingue deux formes d'accentuation⁵ : øjǿw (latin *acutum*, pointu, aigu); barǿw (latin *gravis*, lourd, grave). Comme l'observe Allen (1973 : 230), ces deux types accentuels ne correspondent pas à une simple opposition de hauteur — du type haut *versus* bas —, puisque le type barǿw (grave) y est employé en référence au bruit du tonnerre. Il faut noter en plus que l'accentuation øjǿw (aigu) est positivement marquée dans la mesure où elle ne peut affecter qu'une et une seule syllabe du mot. Pour ce qui est de leur nature, ces *accentuations* seraient des variations de *ton*, assimilables avant tout à des variations mélodiques ou, si l'on veut, à des fluctuations de nature musicale. Les termes grecs utilisés pour les désigner, t'asiw (étirement) et tóyow (ton), renvoient respectivement à la tension des cordes d'un instrument de musique et à la hauteur d'une note de musique (Allen 1968 : 106; 1973 : 230); l'étude des tóyow (ton) relève de la *prosvd'a* (prosodie).

À ces deux modalités de l'accentuation, Aristote, quelques années après Platon, en ajoute une troisième (Collart 1954 : 70; Allen 1968 : 112) : le tóyow (ton), dit m'evs (intermédiaire). Il est cependant bien difficile de préciser ce dont il s'agit; diverses hypothèses ont été proposées : son transitoire (*glide*) qui suivrait un ton øjǿw (aigu), ton composé qui serait lui-même constitué d'un ton øjǿw (aigu) et d'un ton barǿw (grave), etc. (cf. Allen 1968 : 112). Quoi qu'il en soit, et en dépit de la difficulté de le définir, cet intermédiaire est mentionné plus tard par des grammairiens grecs — Glaucos de Samos, Hermocrates d'Iasos, dont on ne connaît plus que les noms (Collart 1954 : 71) — puis, selon le témoignage du Pseudo-Sergius (sur ce grammairien, cf. *infra*), par le grammairien latin Varron au I^{er} siècle avant notre ère (Collart *ibidem*), puis encore, quelques siècles plus tard, par plusieurs grammairiens de la basse latinité et enfin par Jean de Garlande, au XII^e siècle, entre autres.

Constatation capitale : à partir du III^e siècle avant Jésus-Christ, les deux premières modalités accentuelles — øjǿw (aigu) et barǿw (grave) — ont une correspondance à l'écrit. Les savants de l'école alexandrine — Zénodote, Aristarque et surtout Aristophane de Byzance, pour mentionner les plus célèbres — éditent alors les textes d'Homère qui

⁵ Cette distinction reprend peut-être une nomenclature antérieure, celle établie par les phonéticiens de l'Inde ancienne dans leurs études sur la langue védique (Allen 1973 : 231).

dateraient du IX^e siècle avant notre ère. Or, les textes homériques faisaient problème à plusieurs points de vue : une langue ancienne qui présentait des difficultés de compréhension, des frontières de mots ou de groupes de mots difficilement identifiables, des transcriptions vraisemblablement erronées, la présence fréquente d'homographes, etc. (Chantraine 1958 : 5-16; Allen 1968 : 114). Pour faciliter la lecture et la compréhension de ces textes, ces éditeurs ont recours à une série de signes diacritiques (Beaulieux 1927 : 1; Chantraine 1958 : 192; Collart 1968 : 34; Allen 1968 : 114; Allen 1973 : 244). Outre les signes diacritiques liés directement à la phonétique du grec ancien (esprit rude, semblable à notre accent grave, et à l'esprit doux, dont le dessin ressemble à notre apostrophe), cette série de signes graphiques comprenait :

- un signe de «séparation», indiquant les frontières de mots et ressemblant à notre virgule (Allen 1973 : 4);
- un signe de «ligature», sorte de trait d'union arrondi et souscrit (Allen 1973 : 228);
- un signe indiquant la suppression d'une voyelle, qui est visuellement semblable à une apostrophe et qui est parfois difficile à distinguer de l'esprit rude (Allen 1968 : 92; 1973 : 228);
- deux signes qui précisent la longueur, brève ou longue, de certaines voyelles (Allen 1968 : 86);
- et surtout une triade graphique de trois signes qui font songer, respectivement, à l'accent aigu, à l'accent grave et à l'accent circonflexe.

C'est cette série de signes diacritiques que l'on retrouve, recopiée souvent intégralement, chez de nombreux grammairiens latins au IV^e siècle (cf. *infra*).

Il y a lieu de mentionner en outre que l'emploi du terme de *prosvd^a* (prosodie), qui désignait jusqu'alors les *tóyow* (tons), s'est étendu et fait référence dès cette époque à cette série de signes diacritiques (Uhlig 1884 : 170; Allen 1973 : 4). C'est ce que semble suggérer la *TEXNH* de Denys le Thrace⁶ (éd. Uhlig 1884 : 105), puisque ce grammairien y énumère tous ces signes qu'il dénomme *prosvd^{ai}* (prosodies). Comme l'ont remarqué Uhlig (1884 : 170) et Allen (1973 : 4), cette extension sémantique de *prosvd^a* apparaît justifiée si l'on songe que l'utilisation des signes de ligature ou d'apostrophe, par exemple, permettait de savoir si l'on avait affaire à un ou à deux mots, ce qui facilitait sans doute l'accentuation; cette extension sémantique du terme *prosvd^a* (prosodie), pour désigner à la fois des variations de tons et des signes graphiques, conduit cependant à mettre sur le même plan des phénomènes d'ordre phonétique et des faits de graphie⁷.

Qu'en est-il de la triade de signes graphiques proprement dite ? Dès cette époque, ces trois signes sont présentés dans l'ordre habituel; dans sa *TEXNH*, Denys le Thrace (éd. Uhlig 1884 : 105) successivement mentionne et représente l'*øjÿw* (aigu), le *barÿw* (grave) et le *perwspvm`eyh* (circonflexe). Bien que ce grammairien ne l'indique pas, il semble que,

⁶ La plupart des ouvrages qui traitaient de grammaire, avant Varron (I^{er} siècle avant notre ère) et bon nombre des traités de Varron lui-même, ont disparu. La *Techné* attribuée à Denys le Thrace (fin du II^e siècle avant notre ère) constitue une exception (Collart 1954 : 1, 137, 143, 267, 305; Holtz 1982 : 75-78). L'authenticité de ce texte, que l'on croyait acquise, est remise en cause, à tort ou à raison (Holtz 1982 : 78).

⁷ L'utilisation du terme *prosvd^a* (prosodie) dans ce contexte aurait d'ailleurs été considérée comme relevant d'un mauvais usage dès cette époque (Allen, 1973 : 4).

dès le début, l'*accent aigu* indique un ton øjʁw (aigu), l'*accent grave* indique un ton barʁw (grave), au début au moins. et qu'enfin l'*accent circonflexe* indique la première more d'une voyelle longue ou d'une diphtongue (Uhlig 1884 : 6-7; Allen 1968 : 114-115). Tels sont les principes; il s'en faut pour que leur application soit uniforme. Allen (1968 : 115), entre autres, signale que le recours systématique à l'*accent grave* a été évité, parce que ce signe graphique n'était que rarement indispensable; en outre, l'emploi des accents n'est pas toujours constant (Chantraine 1958 : 5-16, Allen 1968 : 114; Allen 1973 : 228). Enfin, ce système de signes diacritiques ne s'est pas généralisé dans la Grèce antique; de nombreux documents l'attestent (Beaulieux 1927 : 1; Allen 1968 : 87).

Pour ce qui est des grammaires de l'époque latine classique, il faut rappeler que l'on ne dispose que des fragments, édités par Funaioli (1907), de la grammaire de Remmius Paléon, qui a vécu au I^{er} siècle avant notre ère (cf. Collart 1938; Holtz 1981 : 77) et des bribes des grammaires de Terentius Scaurus et d'Arrentius Celsus (Holtz 1981 : 77). En dépit des lacunes de notre documentation, il y a lieu de croire cependant que le mot *accentus* a pu être employé dès cette époque, vraisemblablement chez le grammairien latin Varron (116-27, avant notre ère). Vraisemblablement seulement, car reconstituer la notion d'accent pour Varron est un exercice périlleux, puisque le point de vue de ce grammairien n'est accessible qu'à partir des commentaires de grammairiens ayant vécu souvent plusieurs siècles plus tard. La difficulté est d'autant plus grande que le point de vue de Varron sur l'*accentus* semble ne pouvoir être reconstruit qu'à partir des écrits du Pseudo-Sergius (Collart 1954 : 64-65). Or, on ignore pratiquement tout⁸ sur le Pseudo-Sergius; tout au plus peut-on supposer qu'il a vécu entre le V^e et le VII^e siècle de notre ère⁹, soit près de 600 ans après Varron. La transformation du point de vue de Varron a pu, on s'en doute, être altérée par le temps. Il faut croire néanmoins que Varron a dû être un des premiers grammairiens à recourir à la notion d'*accentus*.

Mais si, en dépit de ces réserves, l'on tient compte de ce témoignage, le son articulé de la voix humaine, pour Varron (traduction Collart 1954 : 63) :

«comme tout corps matériel, a trois dimensions : la longueur, la hauteur et l'épaisseur. C'est à l'aide du temps et des syllabes qu'on évalue la longueur car le délai consacré à l'émission des mots et la valeur des syllabes importent énormément. C'est l'accent qui décèle la hauteur [*altitudinem discernit accentus*], selon que telle partie du mot [*cum pars verbi*] [ou] tombe dans le grave [*aut in grave deprimitur*] ou s'élève dans l'aigu [*aut sublimatur in acutum*]. L'épaisseur est fonction du souffle»¹⁰.

⁸ L'existence même de ce grammairien est problématique. Il est possible que ce nom ne soit qu'une déformation du nom du grammairien Servius, dont l'enseignement à Rome a dû débiter dans le courant de l'an 400 de notre ère. Il est possible aussi qu'un grammairien du nom de Servius ait vécu à la même époque que Servius ou à une époque postérieure (Holtz 1981 : 226-227).

⁹ Selon Hubert (1972 : 146), ce grammairien serait «carolingien sans doute»; selon Holtz, il fait partie des grammairiens ayant vécu entre le V^e et le VII^e siècle (Holtz 1977 : 525).

¹⁰ «Scire oportet vocem, sicut omne corpus, tris habere distantias, longitudinem, altitudinem, crassitudinem. Longitudinem tempore ac syllabis metimur; nam et quantum morae enuntiandis verbis teratur et quanto numero modoque syllabarum unum quodque sit verbum, plurimum refert. Altitudinem discernit accentus, cum pars verbi aut in grave deprimitur aut sublimatur in acutum.

Cette théorie des trois propriétés du son — la longueur, la hauteur et l'épaisseur — rappellerait les théories phonétiques des Stoïciens¹¹. En ce qui concerne l'accentuation proprement dite, on relève chez Varron, au I^{er} siècle avant notre ère, d'une part, la correspondance entre la hauteur et l'*accentus*, et, d'autre part, la possibilité de deux hauteurs, selon que telle partie du mot [*cum pars verbi*] [ou] tombe dans le grave [*aut in grave deprimitur*] ou s'élève dans l'aigu [*aut sublimatur in acutum*].

À Rome, la triade aigu, grave et circonflexe est aussi intégrée à la rhétorique. Dans l'*Oratore*, au I^{er} siècle avant notre ère, Cicéron la mentionne dans des considérations sur la voix (traduction et édition, Yon 1964 : 20-21) :

«Remarquable est le caractère de la voix dont avec trois tons en tout, circonflexe, aigu et grave, on obtient une diversité si grande et si suave dans le chant»¹².

Bien que le terme *accentus* ne fasse pas partie du vocabulaire choisi par Cicéron pour désigner la triade, le rapport qu'il établit ici avec le chant importe dans la mesure où l'étymologie ultérieure fait d'*accentus* un dérivé de *ad+cantus* (vers le chant). Au I^{er} siècle de notre ère, Quintilien intitule *De accentibus* un passage de son *Institution oratoire*, passage où il traite des erreurs d'accentuation (traduction et édition Cousin 1975, I, 5 : 92-93) :

«Il est difficile de détecter les erreurs qui concernent les tons, *tenores* [...], ou les accents, en grec *prosvd^a* [prosodie]; ainsi quand l'aigu et le grave sont employés l'un pour l'autre [...] ou la substitution du grave au circonflexe»¹³.

Dans le chapitre qu'il dédie au style, *De genere dicendi*, Quintilien revient, à une autre reprise, sur les trois modalités de l'*accentus* (traduction et édition Cousin 1975, XII, 10 : 123) :

«Mais notre accentuation aussi, à la fois par une certaine rigidité et surtout par sa seule monotonie, est moins agréable, parce que la dernière syllabe n'est jamais relevée par un accent aigu, ni prononcée avec un accent circonflexe, mais le <mot> se termine toujours par une ou deux syllabes à accent grave»¹⁴.

Crassitudo autem in spiritu est» (Keil 1857, IV : 525-526; Funaioli 1907 : 296; cité par Collart 1954 : 63).

¹¹ Pour les Stoïciens, la dialectique commence par une définition du son, élément premier du mot, le mot étant la base du discours. Diogène de Babylonie, Archédème et Antipater de Tarse ont tous écrit un traité concernant les sons; quant à Chrysippe, il en traite dans le II^e livre de sa physique (cf. Collart 1954 : 56-58).

¹² «Mira est enim quaedam natura uocis, cuius quidem e tribus omnino sonis, inflexo, acuto, gravi tanta sit et suavis perfecta in cantibus» (édition, Yon 1964 : 20-21).

¹³ «Adhuc difficilior observatio est per tenores [...] uel accentus, quas Graeci *prosvd^a* uocant, cum acuta et gravis alia pro alia ponuntur [...], si acuitur prima aut gravis pro flexa [...]» (édition Cousin 1975, I, 5 : 92-93).

¹⁴ «Sed accentus quoque, cum rigore quodam, tum similitudine ipsa, minus suaues habemus, quia ultima syllaba nec acuta umquam excitatur nec flexa circumducitur, sed in grauem uel uas gravis cadit semper» (édition Cousin 1975, XII, 10 : 123).

Dans ces trois citations, on retrouve certes la triade accentuelle, mais il n'y est faite aucune mention d'une corrélation entre les trois modalités de la voix et l'écrit. C'est chez les grammairiens de la basse latinité qu'un tel rapport est établi.

Au IV^e siècle de notre ère, Aelius Donatus, intitulé *De tonis* une des sections de son *Ars maior*. Donat commence par mentionner que les termes *accentus*, *tonos* et *tenores* sont équivalents et qu'il y a trois accents : l'aigu, le grave et le circonflexe¹⁵. Donat enchaîne en examinant la place des accents dans le mot. Jusque-là, les considérations du grammairien sont de nature phonétique¹⁶. Sans transition, Donat passe à des considérations sur les signes diacritiques élaborés par les éditeurs alexandrins. Pour Donat, l'accent aigu est une marque qui monte en oblique vers la droite, le grave est une marque qui descend du sommet vers la droite, le circonflexe est une marque composée d'un aigu et d'un grave¹⁷. C'est dire que sous la rubrique de l'accent, les modalités de l'accentuation et les signes auxiliaires de l'alphabet sont réunis. Donat rapproche la voix et la main.

À la même époque, dans son *Artis grammaticæ libri III* (Keil 1857-1880 : I, 430; Hubert 1970 : 32), le grammairien Diomède tient les mêmes propos et définit les trois accents dans les mêmes termes¹⁸ que Donat¹⁹. Un peu plus tard, au VI^e siècle, dans un texte attribué à Priscien, les *Partitiones duodecim uersuum Æneidos principalium* (Keil 1857-1880 : III, 459; Hubert 1970 : 60), l'accent aigu «rend aiguë ou élève la syllabe» (*quod acuat siue euelet syllabam*); l'accent grave l'«abaisse ou la rabaisse» (*quod deprimat aut deponat*); l'accent circonflexe l'«abaisse et l'élève» (*quod deprimat et acuat*). Si l'on considère le verbe *erigat* (de *erigere*, relever) comme un synonyme d'*euelet* (de *euelevare*, lever, élever), ce sont, pour les trois accents, des définitions identiques que retiennent les auteurs suivants : Isidore de Séville, au VII^e siècle (*Etymologarium siue originum*, Hubert 1970 : 69); Hughes de Saint-Victor, au XII^e siècle (*Opera propædeutica. De Grammatica, idem* : 105); Vincent de Beauvais, au XIII^e siècle (*Speculum Doctrinale, idem* : 143). À partir du XII^e siècle, selon les fragments édités par Thurot (1869 : 393), l'expression «élévation de la voix» (*elevatio vocis*) apparaît chez Pierre Hélie²⁰ pour désigner l'accent aigu; «abaissement de la voix» (*depressio vocis*) désigne l'accent grave. Ces définitions, ce sont pratiquement les définitions de Nicot.

¹⁵ «Tonos alii accentus, alii tenores nominant. Toni igitur tres sunt, acutus, gravis, circumflexus» (Holtz 1981 : 609).

¹⁶ Ce faisant, Donat, comme plusieurs grammairiens latins de la basse latinité, aurait appliqué les principes de l'accentuation grecque à l'accentuation latine, sans tenir compte des différences qui existeraient du point de vue de l'accentuation entre ces deux langues (Allen 1973 : 151).

¹⁷ «Acutus accentus est nota per obliquum ascendens in dexteram partem /, gravis nota a summo in dexteram partem descendens \, circumflexus nota de acuto et graui facta» (Holtz 1981 : 610).

¹⁸ «Acutus accentus est nota per obliquum ascendens in dexteram partem /, gravis nota a summo in dexteram partem descendens \, circumflexus nota de acuto et graui figuratur» (Hubert 1970 : 32; Keil 1857-1880 : I, 430).

¹⁹ Cet accord entre les définitions de ces deux grammairiens n'est pas inhabituel (Holtz 1982 : 84-85)..

²⁰ Sur les questions d'éditions reliées à l'oeuvre de P. Hélie, voir Lusignan (1986), entre autres.

Au terme de ce parcours, il me semble possible de tirer quelques conclusions. La première concerne la définition de l'accent comme «élévation de la voix». Si le *Dictionnaire historique de la langue française* situe l'émergence de cette définition en 1559 (Rey 1995 : I), il y a tout lieu de croire que des définitions voisines avaient été proposées en latin pour l'*accentus* avant cette date. Selon toute vraisemblance, choisir de définir l'accent comme une «élévation de la voix», c'est sans doute retenir, pour l'accent, envisagé globalement, la vieille définition du seul accent aigu.

J'ajouterai, à propos de cette définition de l'accent comme une «élévation de la voix», et de façon à rappeler les limites de cette enquête, que de nombreux autres textes pourraient bien sûr être étudiés de près. L'examen de la *Grammaire générale et raisonnée* d'Arnaud et Lancelot (1660 : 16) montrerait que l'accent y est aussi défini comme une «élévation de la voix»; pour Arnaud et Lancelot «l'élévation de la voix s'appelle accent *aigu*, et le rabaissement, *accent grave*» (*idem*). Une étude détaillée des grammaires²¹, des dictionnaires et des encyclopédies, des traités de versification, des manuels de rhétorique et de musique, montrerait des affinités et mettrait à jour des divergences, bien sûr.

Enfin, ce que ces pages illustrent sans doute, c'est la prégnance à travers l'histoire de cette analogie qui associe les trois signes diacritiques avec les trois modalités phonétiques de l'accentuation, cette analogie qui, avec ses nuances, s'observe du IV^e au XVIII^e siècle.

RÉFÉRENCES

- ACADÉMIE (1932) : *Dictionnaire de l'Académie française*, Paris, Hachette (8^e édition); 1878, *idem*, Paris, Didot (7^e édition); 1835, *idem* (6^e édition); 1798, *idem*, Paris, Smits (5^e édition); 1762, Paris, Brunet (4^e édition); 1740, *idem*, Paris, Coignard (3^e édition); 1694, *idem*, Paris, Coignard (1^{re} édition) [Reproduction AUPELF-UREF].
- ALLEN, W.S. (1968) : *Vox Graeca*, Cambridge, University Press.
- ALLEN, W.S. (1973) : *Accent and Rhythm*, Cambridge, Cambridge University Press.
- ARNAUD, A. et Cl. LANCELOT (1660) : *Grammaire générale et raisonnée*, Paris, Reproductions Paulet.
- AUGÉ, C. (1898) : *Nouveau Larousse illustré*, Paris, Larousse.
- BEAULIEUX, Ch. (1927) : *Histoire de l'orthographe française. Tome second. Les Accents et autres signes auxiliaires*, Paris, Champion.
- BESCHERELLE, A. (1893) : *Dictionnaire national*, Paris, Garnier; (1874) : *idem*; (1867) : *idem*.
- CARTON, F. (1974) : *Introduction à la phonétique du français*, Paris et Montréal, Bordas.

²¹ Selon Stengel (1890), avant 1800, 625 grammaires ont été publiées; à ces 625 grammaires, il faudrait ajouter les 2037 grammaires qui, selon Chervel (1982, p. XIII) ont vu le jour de 1800 à 1914, et à cela s'ajouteraient toutes celles qui ont été imprimées depuis.

- CHANTRAINE, P. (1958) : *Grammaire homérique*, Paris, Klincksieck.
- CHERVEL, A. (1982) : *Les Grammaires françaises 1800-1914*, Paris, Institut national de recherche pédagogique, Service d'histoire de l'éducation.
- CLAS, A. (1983) : *Sons et langage*, Montréal, Sodilis.
- COLLART, J. (1938) : «Palémon et l'*Ars Grammatica*», *Revue de philologie*, XII, pp. 228-238;
- COLLART, J. (1954) : *Varron grammairien latin*, Paris, Les Belles Lettres.
- COUSIN, J. (1964) : *Quintilien. Institution oratoire*, texte établi et traduit par J. Cousin. Paris, Les Belles Lettres.
- DAUZAT, A., DUBOIS, J. et H. MITTERAND (1964) : *Dictionnaire étymologique et historique du français*, Paris, Larousse (réédition, 1994).
- FUNAIOLI, H. (1907) : *Grammaticae Romanae Fragmenta*, Leipzig, Teubner (réimpression, Stuttgart, Teubner, 1969).
- GARDE, P. (1968) : *L'Accent*, Paris, PUF.
- HATZFELD, A. et A. DARMESTETER (1924) : *Traité de la formation de la langue*, Paris, Delagrave (réimpression intégrale, Paris, Delagrave, 1964).
- HOLTZ, L. (1977) : «À l'école de Donat, de saint Augustin à Bède», *Latomus*, XXXVI (2), pp. 522-537.
- HOLTZ, L. (1981) : *Donat et la tradition de l'enseignement grammatical*, Paris, CNRS.
- HUBERT, M. (1970) : «Corpus Stigmatologicum Minus», *Archivum Latinatis Medii Aevi*, XXXVII, pp. 6-171.
- HUBERT, M. (1972-1) : «Le vocabulaire de la "ponctuation" aux temps médiévaux, un cas d'incertitude lexicale», *Archivum Latinatis Medii Aevi*, XXXVIII, pp. 57-169.
- HUBERT, M. (1972-2) : «Isidore de Séville novateur ?», *Revue des études latines*, 1971 (49^e année), pp. 290-319.
- KEIL, H. (1857-1880) : *Grammatici Latini*, Leipzig, Teubner (réimpression, 1961, Hildesheim, Olms).
- LAROUSSE, P. (1866) : *Grand Dictionnaire universel du XIX^e siècle*, Paris, Larousse et Boyer.
- LÉON, P. (1992) : *Phonétisme et prononciations du français*, Paris, Nathan.
- LITTRÉ, É. (1873) : *Dictionnaire de la langue française*, Paris, Hachette; (1885) : *idem*.
- LUSIGNAN, S. (1986) : *Parler vulgairement*, Paris et Montréal, Vrin et Presses de l'Université de Montréal.
- MÉNAGE, G. (1644) : *Dictionnaire étymologique ou Origines de la langue françoise*, Paris, Anisson [Reproduction AUPELF-UREF].

- NICOT, J. (1621) : *Thresor de la langue françoise*, Paris, Douceur (réimpression, Paris, Picard, 1960).
- OLIVET, Abbé d' (1753) : *Traité de prosodie française*, Genève, Frères Cramer [Reproduction AUPELF-UREF].
- LOUDON, A. (1640) : *Curiositez françoises*, Paris, Sommaville [Reproduction AUPELF-UREF].
- QUEMADA, B. (1967) : *Les Dictionnaires du français moderne 1539-1863*, Paris, Didier.
- REY, A. (1989) : *Dictionnaire alphabétique et analogique de la langue française*, Paris, Société du Nouveau Littéré - Le Robert.
- REY, A. (1992) : *Dictionnaire historique de la langue française*, sous la direction de Alain Rey, Dictionnaires Le Robert (nouvelle édition, 1993).
- RICHELET, P. (1680) : *Dictionnaire françois contenant les mots et les choses*, Genève, Widerhold.
- RICHELET, P. (1681) : *Dictionnaire de la langue française ancienne et moderne*, Lyon, réédition à partir de l'original, Duplain, 1759; (1732) : *idem*, Amsterdam, Aux dépens de la compagnie.
- ROBERT, P. (1971) : *Dictionnaire alphabétique et analogique de la langue française*, Paris et Casablanca, Société du Nouveau Littéré; (1957) : *idem*.
- ROSSI, M. (1979) : «Le français, langue sans accent», pp. 24-47 in I. Fonagy et P. Léon, *L'Accent en français contemporain*, Ottawa, Didier.
- ROSSI, M., et A. DI CRISTO, D. HIRST, Ph. MARTIN et Y. NISHINUMA (1981) : *L'Intonation. De l'acoustique à la sémantique*, Paris, Klincksieck.
- STENGEL, E. (1890) : *Chronologisches Verzeichnis französischer Grammatiken vom Ende des 14. bis zum Ausgange des 18. Jahrhunderts nebs Angabe des bisher ermittelten Fundorte derselben*, Oppeln.
- THUROT, C. (1869) : *Extraits de divers manuscrits latins*, Paris (réimpression, Frankfurt, Minerva, 1964).
- TOBLER, A. et E. LOMMATSCH (1925 ff) : *Altfranzösisches Wörterbuch*, Berlin.
- UHLIG, G. (Ed.) (1884) : *Dionysius Thrax (ca. a. 100 a. C.), Techne*, Grammatici Graeci I, I, Leipzig, Teubner.
- YON, A. (1964) : *Cicéron. L'orateur. Du meilleur genre d'orateur*, texte établi et traduit par A. Yon, Paris, Les Belles Lettres.

«MOTIVATION DÉMOTIVATION REMOTIVATION» OU LE CYCLE DES MOTS

Blandine PÉRICHON

Université de Provence-Aix-Marseille, France

Nous nous proposons, à l'aide de quelques exemples extraits d'un corpus se rattachant aux domaines de l'écologie et de l'environnement en France, d'analyser comment l'élargissement de la sphère d'énonciation de ce discours, avec l'extension de l'écologie du champ scientifique au champ socio-politique, a pu et peut avoir des influences sur le sémantisme des unités lexicales en extension comme en compréhension.

Nous allons nous pencher sur les questions de motivation et/ou de démotivation des termes en observant la manière dont chaque groupe énonciatif, au sein d'un domaine assez vaste que nous nommons «de l'écologie et l'environnement», déploie des stratégies énonciatives et terminologiques propres.

Nous verrons comment s'effectuent des opérations «d'appropriation» de termes qui apparaissent, dans certains cas avec une resémantisation du signifié, où qui peuvent aller jusqu'à la création d'un nouveau terme. Nous nous intéresserons alors aux questions de «négociation de la dénomination» (Gaudin 1993 : 191). Nous observerons également un phénomène de «mise à l'écart» de lexèmes qui vont être vécus comme des éléments marqués par les autres groupes (du type «comme ils disent»).

PRÉSENTATION DU DOMAINE D'ÉTUDE ET PRÉCISIONS MÉTHODOLOGIQUES

L'écologie est une science ancienne de plus d'un siècle. Le terme *écologie* a été créé en 1866 par l'allemand E. Haeckel (formé à partir du grec *logos* : «discours», et de *oikos* : «maison, habitat»). À la fin des années soixante un mouvement social s'amorce et l'écologie devient une des composantes des mouvements alternatifs et contestataires. Les années soixante-dix vont voir l'émergence d'associations de protection de la nature et des luttes antinucléaires. L'écologie en tant que mouvement social et politique se développe indépendamment mais souvent en parallèle avec la science dans la mesure où elle reprend des questions soulevées par les scientifiques.

Notre domaine d'application que nous nommons «de l'écologie et l'environnement» s'étale sur une période de vingt ans (de 1974 à 1994). Les exemples que nous prendrons seront extraits de notre corpus de thèse qui est composé de quatre revues qui couvrent

différents pans de ce vaste domaine; de l'écologie scientifique appliquée, technique jusqu'à l'écologie politique et théorique.

Notre corpus est composé de la revue *Aménagement et Nature* qui est une revue trimestrielle (qui existe depuis 1966), que l'on qualifiera de scientifique appliquée et technique, destinée à un public averti mais non spécialiste (techniciens, élus locaux, chercheurs, ...). Nous avons également la revue *Combat Nature* qui se présente comme «la revue des associations écologiques et de défense de l'environnement» (dans son sous-titre). Cette revue trimestrielle, reflète donc l'écologie associative militante depuis 1971. Ensuite une revue qui couvre le champ de l'écologie politique partisane, *Vert-Contact* qui est un feuillet hebdomadaire (entre une et deux pages) dont le premier numéro date de septembre 1986. Ce feuillet hebdomadaire s'il ne se présente pas comme «la lettre interne des Verts» est toutefois réalisé par leurs soins et destiné à un public averti et partisan. La dernière revue, la plus récente, est *Écologie politique* (depuis 1991) qui comprend une centaine de pages par numéro et propose une réflexion théorique sur l'écologie et l'environnement. Le responsable de la rédaction, Jean-Paul Deléage est un militant écologiste avec une formation d'écologue.

Notre corpus est constitué à partir d'un relevé de mots «pivots» (environ 150 mots)¹ que nous avons relevé de façon exhaustive (l'intégralité de chaque numéro sélectionné); ce qui constitue un corpus conséquent (800 pages). Ces mots-pivots sont relevés avec leur contexte énonciatif (titre article, nom d'auteurs, ainsi que la ou les phrases qui les entourent afin de permettre une compréhension acceptable de ce contexte).

Nous allons prendre quelques exemples de ce vocabulaire dans notre corpus afin d'observer le processus de motivation démotivation remotivation du sens d'un ensemble de termes sélectionnés.

Nous précisons tout d'abord que par *motivation* nous entendons :

«l'ensemble des facteurs conscients ou semi-conscients qui conduisent un individu ou un groupe à avoir un comportement déterminé dans les actes de parole».²

Dans le processus de créativité lexicale et de dérivation qui nous intéresse, on parlera de motivation pour définir la relation entre le mot (ou le préfixe) et le dérivé.

LA MOTIVATION DES TERMES

Nous allons nous pencher sur quelques exemples, et notamment sur des termes composés par les préfixes *éco-* et *bio-*.

Dans l'état actuel de notre recherche, nous parlerons de *préfixe* pour désigner les éléments de composition (ou dérivation) que sont *éco-* et *bio-*. Cette notion peut être

¹ Auxquels il faut ajouter les différents syntagmes formés à partir d'un même mot. Par exemple *durable* : *développement durable*, puis *politique durable*, *ville durable*, *environnement durable*, etc.

² 1993, *Dictionnaire de linguistique et des sciences du langage*, Larousse, p. 313.

remise en cause. On pourrait alors parler de *parapraxème* si l'on se situe selon une approche praxémique; ou bien encore de *recomposé moderne* selon Martinet et d'autres notions pourraient être avancées. Nous conserverons pour l'heure la notion de préfixe dans la mesure où l'élément *éco-* tout comme l'élément *bio-* fonctionnent bien comme des morphèmes antéposés à une base qui permettent la création de mots nouveaux.

Le préfixe *éco-*

Dans les années soixante-dix le préfixe *éco-* (du grec *oikos* «maison, habitat») se généralise et d'un point de vue sémantique il est le fruit de la combinaison des sémantèmes issus par troncation de *économie* et *écologie*. On parle alors d'*éco-développement*, *éco-technologie*, *éco-habitat* ou encore d'*éco-énergies*, ou d'*éco-produit*.

- **L'écotechnique** servant de base à un **éco-développement**. (*Aménagement et Nature*, 1976 : 69)
- Il faut en contrepartie un **éco-développement et construire une éco-société** qui soit appropriée à une vie saine de l'homme et de la planète. (*Combat Nature*, 1978)
- **L'écohabitat** n'existe pas, mais on peut y tendre; on peut faire en sorte que l'habitat soit moins brutaliste vis-à-vis **des milieux**. (*Aménagement et Nature*, 1976)
- L'occupation par **les groupes écoculturels des niches disponibles dans les écosystèmes humaines est le fait tant d'impératifs biologiques que d'impératifs culturels**. (*Aménagement et Nature*, 1975 : 11)

Le préfixe *éco* non autonome est un préfixe très productif dans notre corpus global (beaucoup plus que *bio-*, comme nous le verrons par la suite). Il sert à la construction d'une cinquantaine de mots ou dérivés (*écologie*, *écologiste*, *écologiquement*, *écohabitat*, *écocrate*, *éco-impôt*, *écotaxe*, etc.) que nous avons répertoriés 5933 fois dans notre corpus total. Parmi ces différents dérivés nous comptabilisons 125 occurrences qui contiennent *éco-* marqué par un tiret, comme :

- Une ambition : développer **l'éco-citoyenneté**. (*Combat Nature*, 1994)
- ... la création de la norme NF Environnement, la diffusion **des éco-bilans** dans les entreprises, la réforme de la politique agricole commune, la relance de la politique de délocalisation. (*Combat Nature*, 1992 : 19)
- Et **les protecteurs de la nature** ont encore bien des griefs à adresser non seulement aux industriels ou aux collectivités locales qui se hâtent lentement pour prendre les mesures adéquates, mais aussi au gouvernement dont on souligne la politique trop molle tant en matière réglementaire que dans le domaine de **l'aide aux éco-industries**. (*Combat Nature*, 1986 : 44)

Le sémantisme du préfixe *éco-* se modifie d'un point de vue diachronique. Aujourd'hui il est possible de dire que le sens d'*économie* contenu dans ce préfixe tend à disparaître ou tout au moins a une productivité lexicale beaucoup plus faible que celle rattachée au domaine de l'écologie. Nous pouvons dire plus précisément que ce sens s'estompe et qu'il est désormais, beaucoup moins présent, ou tout au moins de façon explicite.

Dans le domaine environnemental, au cours des années quatre-vingts de nouveaux termes sont créés à la suite de réglementations nouvelles. On parle alors d'*écotaxe* dans le domaine juridique Européen, il est également question d'*éco-audit* ou d'*éco-label*.

- **Une écotaxe**, c'est une taxe qui fait payer les pollueurs. Elle a principalement pour but de dissuader de polluer.³
- **L'éco-audit** : la proposition de règlement communautaire, objectifs et aperçu de ses principaux aspects. Un domaine en pleine évolution : les audits d'environnement. (*Aménagement et Nature*, 1991)
- **Un éco-label** (ou **label vert**) est une marque écologique officielle qu'une autorité publique attribue à certains produits, de qualité plus écologique que les autres de même catégorie. (*Combat Nature*, 1990)

Toutes ces réalisations nous montrent une grande productivité et une motivation forte du préfixe *éco-* qui est là comme marqueur d'un discours qui se veut avant toute considération sémantique «éco», c'est-à-dire qui a trait à l'écologie et l'environnement.

L'intégration des préfixes aux différents termes évoqués est un élément intéressant à observer. Il est possible de distinguer certaines réalisations où le préfixe et le lexème-base sont liés par un tiret, et d'autres au contraire où le préfixe et la base ne «font qu'un», sans trace du processus de dérivation de ce terme.

Nous pensons pouvoir dire après une première observation que ces deux phases allaient se distinguer d'un point de vue diachronique, marquant alors l'intégration du terme dans le discours commun. En fait, au vu de différents exemples, il paraît difficile de discerner une évolution stricte même si l'on peut dire simplement que, plus un terme est nouveau dans le domaine où il est employé, plus il a de chances de contenir un tiret entre le préfixe et le lexème.

On peut nuancer cette évolution suite à l'étude de Dubois et Guilbert (1961) qui montre que la non-intégration totale du préfixe au mot avec la conservation du tiret peut servir de marqueur. Ceci permet alors au préfixe de conserver un sens propre et d'attirer l'attention, ici sur le préfixe *éco-* plus que sur le contenu du lexème. Les Verts parlent par exemple de :

«**l'éco-technologie** est une technologie propre qui, au terme du processus de production, fabrique des produits propres.»

³*Dictionnaire des Verts*, 1994, p. 73.

Ainsi peu importe la technologie ou le consommateur, l'information essentielle c'est que ceux-ci soient «éco», avec toutes les significations que l'on pourra alors donner à ce préfixe.

Le sociologue, H.P. Jeudy aborde la question de la prolifération des préfixes ainsi que celle de leur rôle idéologique. Il pense que «l'usage des préfixes créeraient un réseau de "marques" antérieur à la parole et par lequel se réalisent toutes les différences subjectives»⁴. Ainsi un terme peut apparaître comme technique alors qu'il est politique. Le vocabulaire de l'écologie politique semble alors ainsi marqué. Nous pouvons prendre l'exemple du programme des Verts pour les élections européennes de 1989 qui était basé sur la notion d'*éco-développement*; même si l'on ne connaît pas le contenu de ce «développement», il est déjà qualifié avec l'emploi du préfixe *éco-*. À chacun ensuite de reconstruire le sémantisme de cet *éco-développement*.

Le préfixe *bio-*

Bio- est un élément préfixal, qui est un emprunt au grec *bios-* «vie», qui s'est développé avec l'apparition du mot *biologie* (1802).

En écologie politique on ne trouve actuellement qu'un nombre restreint de termes qui contiennent ce préfixe. Le préfixe *bio-* et les termes qui en sont constitués sont encore fortement marqués par une connotation scientifique; ceci est peut-être le pendant d'une création néologique qui ne s'est pas (ou pas encore) développée.

Le préfixe *bio*, sous la réalisation d'un morphème non autonome est présent dans 573 occurrences, c'est-à-dire qu'il entre dans la composition de termes comme *biologie*, *biologiques*, *biocénose*, *biosphère*, ou *biodiversité*... soit une vingtaine de mots différents.

- La protection de la **biodiversité** : inscrite dans les conclusions du sommet de Rio, la **biodiversité** concerne aussi l'agriculture avec les races animales et les espèces végétales en voie de disparition. (*Vert Contact*, 1994)
- France Nature Environnement reste réservée sur la fabrication des **plastiques biodégradables** et ceci pour trois raisons : - la **biodégradabilité de produits** comme Bio-D reste encore incertaine et la **notion même de biodégradation** serait à remplacer par celle de détérioration; - le **plastique biodégradable n'est pas un emballage écologique**; résout-il le problème de la surconsommation de certains types de plastique (emballages) ? - ne risque-t-il pas de banaliser le **droit de polluer sans se soucier de recycler** ? Ceci va à l'encontre des perspectives d'utilisation de produits à longue durée de vie. (*Combat Nature*, 1990 : 65)
- Il semble qu'on s'apprête maintenant à établir un nouveau cadre juridique pour l'**industrie biotechnologique** qui permettra de breveter des «formes de vie», ce qui inclut les ressources agricoles. (*Écologie Politique*, 1993 : 62)

⁴Cité par BAGGIONI, Daniel (1977) : «Aspects sociolinguistiques de la néologie lexicale : le vocabulaire 'à connotations scientifiques' et le 'fétichisme des mots'», *LAB 17*, Leipzig, pp. 33-40, p. 37.

Le préfixe est marqué comme élément distinct de la base à laquelle il se rattache (avec la présence d'un tiret) dans seulement 24 occurrences du corpus global. En outre on ne relève dans l'ensemble de notre corpus que 6 réalisations où *bio* fonctionne comme un morphème libre, autonome, soit une utilisation des plus restreinte.

- D'ATD Quart-Monde **aux agriculteurs bio** en passant par les associations d'usagers de la santé, des associations de protection de la nature... (*Vert Contact*, 1986)
- 500F pour l'achèvement du projet de **maraîchage bio** à Orx (Landes). (*Vert Contact*, 1989)
- **La chimie, désormais bio**, la génétique, ouvrent la voie **aux «bio industries»**... (*Combat Nature*, 1986)

Un autre exemple de terme avec le préfixe *bio-* est *biocarburant*. L'origine de ce terme est clairement scientifique. Il est défini comme un «carburant obtenu du traitement de la biomasse végétale». *Biocarburant* est employé dans le discours général mais il est en concurrence avec la locution donnée comme son synonyme : *carburant vert*. (On le trouve notamment dans une publicité de *Shell* pour ses *carburants verts* : l'essence sans plomb et le gasoil propre.)

Nous pensons que *vert* est aujourd'hui clairement identifiable par un locuteur non spécialiste, et renvoie à un référent lié à l'environnement. Il est alors plus évocateur et signifiant pour le consommateur la même idée que le préfixe *bio-*. La fonction scientifique de *bio-* (renvoyant dans le cas de *biocarburant* à biomasse et non à biologique) de par sa formation sémantique n'est pas nécessairement perçue par le destinataire. En passant à *carburant vert* la motivation scientifique disparaît, tandis qu'une nouvelle motivation «grand public» apparaît dans le syntagme.

Le phénomène est identique avec *éco-produit* que nous avons pu voir ci-dessus qui a comme équivalent *produit vert*. *Vert* vient alors à la place de *bio-* ou de *éco-*. Nous pouvons peut-être parler de «reformulant» pour reprendre une tournure de Daniel Jacobi. *Vert* est alors employé comme reformulant de termes scientifiques, et marque alors le passage du discours scientifique au discours général. Il permet alors aux lecteurs de «prendre appui sur leurs conceptions antérieures : il offre une valeur d'ancrage» (Jacobi 1990 : 114). Cette valeur d'ancrage va permettre au locuteur, quel que soit son référentiel discursif (discours scientifique, discours technique, discours public...), de conserver une motivation aux termes employés.

DÉMOTIVATION DES TERMES

Une perte de signification de certains termes peut s'opérer par exemple par une démotivation qui se traduit par une sorte de banalisation de l'élément préfixal qui les constitue.

Si l'on se situe toujours avec des exemples de termes qui contiennent les préfixes *éco-* ou *bio-* nous pouvons observer un phénomène de démotivation de ces derniers, dans la mesure où ils ont une productivité très forte.

Jean Peytard⁵ pense que :

«la connexion du préfixal avec des bases multiples non seulement élargit l'usage, mais tend de surcroît à grammaticaliser le procédé et à enrichir le contenu sémantique du préfixal».

Nous pensons et cela est indéniable, que l'extension d'emploi d'un élément préfixal à des bases diverses tend à grammaticaliser effectivement le procédé. Pour ce qui est du contenu sémantique du préfixe il nous semble que le phénomène se décline diachroniquement, en plusieurs phases. Le contenu sémantique peut, comme J. Peytard le souligne, être enrichi dans un premier temps, mais cet enrichissement se fait dans le sens d'une pluralité, d'une polysémie qui peut, si elle se développe trop, devenir «excessive», ce qui pourra tendre alors à une certaine neutralisation de son sens. Françoise Madray⁶ précise que «le risque existe, à vouloir faire tout dire à un outil de sens, de le rendre inefficace à signifier quelque chose». Ainsi le préfixe se généralise et perd de sa pertinence. Pour le cas *d'éco-*, son sens pourrait se résumer comme suit : «qui a trait à l'écologie et l'environnement» que chacun va reconstruire approximativement selon le lexème qui s'y rattache, et la situation d'énonciation.

Les médias grand public et la publicité s'emparent alors de cette écologie prometteuse et poursuivent le processus de création néologique; qui pour vanter les mérites d'une *lessive écologique* qui protège l'environnement, qui pour des *éco-emballages* qui ne polluent pas, ou encore comme certains produits de cosmétique (produits *Revlon* et *Clarins*, notamment) qui proposent une crème dénommée «Écologie»⁷(notons également ici le passage du nom commun au nom propre). Le sémantisme des termes en est alors globalement modifié dans la mesure où ils recouvrent des champs de significations multiples pour un même signifiant. Le besoin se fait alors sentir, chez certains énonciateurs, de remotiver des termes qui sont devenus «usés» d'une si grande polysémie.

Dans la mesure où un terme a échappé au domaine duquel il est issu, les locuteurs de ce domaine ont tendance à «renier» la nouvelle acception, pouvant aller jusqu'au rejet du terme lui-même.

En 1987 un écologue, Patrick Blandin⁸, a réalisé une étude sur le concept *d'environnement*. Il s'interrogeait alors sur l'évolution du terme *environnement* qu'on trouve d'une part dans les travaux d'écologie fondamentale où *environnement* désigne «un milieu en tant qu'espace écologique particulier, qu'il s'agisse d'un vaste domaine ou d'un habitat restreint», et d'autre part, dans des titres d'ouvrages et dans des revues de créations récentes qui en font une utilisation fréquente, auquel cas «il s'agit pratiquement toujours,

⁵PEYTARD, Jean (1973) : «De la diffusion d'un élément préfixal : mini-», *Langue française* 17, p. 24.

⁶MADRAY, Françoise (1983) : «Circulation et réglage de sens», *Cahiers de Praxématique* 1, Montpellier, Univ. Paul Valéry, pp. 45-53.

⁷Une crème dont la notice précise qu'elle permet le maintien d'un «taux d'hydratation optimal tout en protégeant des toxines de l'environnement» : l'objectif n'étant pas ici de protéger l'environnement mais de s'en protéger...

⁸BLANDIN, Patrick (1987) : «Le concept d'environnement en recherche en écologie»; Muséum d'histoire naturelle de Paris, non publié.

explicitement ou non, de l'environnement de l'homme considéré le plus souvent dans sa globalité».

On se rend compte alors que les scientifiques souffrent de l'accroissement et de la dérive des termes qui sont issus de leur domaine et se développent à l'extérieur, indépendamment de leur discipline. Au delà de la seule interrogation terminologique, Blandin en vient à remettre en cause son propre champ de recherche et à se demander s'il n'y aurait pas «un véritable concept d'environnement, différent de celui de milieu, qui aurait une valeur heuristique propre dans le champ de l'écologie ?»

UNE PERTE DE SIGNIFICATION QUI ENTRAÎNE UNE REMOTIVATION DES TERMES

Aux yeux de certains énonciateurs, quelques termes semblent perdre de leur pertinence d'un point de vue sémantique, et de ce fait ils ajoutent un autre élément (préfixe ou autre lexème) qui va permettre un nouveau marquage de sens. C'est ainsi que dans notre corpus nous trouvons les énoncés suivants :

- Cela implique-t-il que **l'écologue scientifique** et le biologiste doivent être les conseillers du prince ou cela implique-t-il un autre rapport du politique et du scientifique ? (*Écologie Politique*, 1993 : 8)
- En effet, les emplois qui ne sont pas **écologiquement soutenables** ne peuvent pas être des emplois stables dans le futur. (*Écologie Politique*, 1993 : 129)

Le mot *écologue* ne semble plus sous-entendre un «spécialiste d'écologie» ou un «scientifique qui étudie l'écologie» (vs un écologiste : «défenseur de la nature», distinction faite aujourd'hui par les différents dictionnaires généralistes et spécialisés) car l'énonciateur se sent obligé de préciser qu'il est question d'*écologue scientifique* (vs *écologue militant*, comme par exemple Antoine Waechter qui a une formation d'écologue tout en étant un des leaders de l'écologie politique française). Nous pouvons parler d'un discours marqué par une redondance polémique.

Le phénomène de démotivation/remotivation peut donc se manifester par l'ajout d'un élément (préfixe ou lexème) qui, au delà d'une fonction de marquage supplémentaire dans l'intention de resémantiser un signifié édulcoré par l'usage, ne semble pas réellement porteur d'un sémantisme nouveau et spécifique.

- [...] **la lutte pour un développement écologiquement soutenable** est très étroite. (*Écologie Politique*, 1993 : 131)
- La population n'accepte plus n'importe quel aménagement et d'ailleurs, la France, qui a signé les accords de Rio, reconnaît la nécessité de mettre en place **un développement durable et soutenable**. (*Combat Nature*, 1994 : 9)
- 3^{ème} Fête de l'Homme, la Nature et l'Environnement. Foire régionale du pain, du vin et **fromage éco-biologique**, samedi 19 et dimanche 20 septembre 87 à Rotonde. (*Vert Contact*, 1987 : 2)

Il y a une duplicité sémantique, ainsi au lieu de parler d'un *fromage biologique*, il est alors question d'un *fromage éco-biologique*, si tant est qu'il y ait une différence entre les deux fromages.

Le phénomène peut également s'opérer uniquement sur un signifié, le signifiant restant le même. Nous pouvons citer l'exemple d'*écocide*, terme issue de l'écologie scientifique, qui désigne «une destruction totale des milieux naturels». L'écologie politique reprend ce terme en jouant et insistant clairement sur l'idée de destruction (avec le suffixe *-cide*) et le fait entrer dans un paradigme «alarmiste et catastrophiste» qui existe déjà dans l'écologie politique partisane.

«MISE À L'ÉCART» DE LEXÈMES

Les différents protagonistes de l'écologie et de l'environnement qui participent à ce processus de création néologique et de lexicalisation de termes, se positionnent d'un point de vue discursif en tant qu'énonciateurs liés à un groupe énonciatif spécifique. Ils s'inscrivent alors en tant que communauté linguistique («nous») et procèdent ainsi à des mises à l'écart de lexèmes du type «comme ils disent» (sous-entendu «les autres»).

- Quantifions ce qui est quantifiable avec le maximum d'objectivité possible; mais, pour la qualité de la vie, respectons une certaine manière «d'être et de penser», **appelée par certains «la conscience écologique»**. (*Aménagement et Nature*, 1973 : 18)

La distinction entre *écologiste* et *écologue* est venue des scientifiques en réaction à l'ambiguïté qui était en train de se développer entre des scientifiques et des personnes qui défendent les valeurs d'une idéologie, d'un mouvement politique. Certains proposent alors d'employer les termes déjà existant «à leur juste valeur» comme *Vert*, qui figure ainsi dans un des plus récent dictionnaire spécialisé⁹ à l'entrée *Verts* (au pluriel). Il est alors intéressant de noter la position du scientifique, vis-à-vis de ce nouveau terme qui est défini ainsi :

«Terme utilisé depuis les années 70 pour désigner un courant d'opinion politique qui s'est constitué dans la totalité des pays de l'Europe occidentale. [...] À l'exception de la France, où ils se sont dénommés écologistes, le terme de Verts est systématiquement utilisé pour les désigner dans la totalité des autres pays d'Europe.»

Il poursuit la définition et précise :

«Ce dernier présente l'avantage d'éviter toute confusion entre science et idéologie politique d'une part, et entre science et environnement d'autre part. Il serait en particulier difficile, au plan sémantique, d'attribuer l'apanage du terme d'écologie, science biologique, à un courant de pensée politique quel qu'il soit. Oserait-on parler de physique politique ou de chimie politique ?»

⁹ RAMADE, François (1993) : *Dictionnaire encyclopédique de l'écologie et des sciences de l'environnement*.

La prise de position semble claire de la part de cet écologue; et son souci de monosémie des termes (*écologiste* vs *vert*) évident.

La question de la «négociation de la dénomination» (Gaudin 1993 : 191) au sein d'un groupe linguistique (communauté scientifique, parti politique, revue éco-philosophique) se pose alors. Il peut y avoir concertation (ou non) quant à la dénomination d'un certain nombre de concepts ou d'idéaux que l'on souhaite véhiculer, transmettre.

Par exemple *Les Verts* français ont publié en 1994 un dictionnaire de l'écologie politique¹⁰. Ce livre marque un besoin fort de dénomination commune au sein du parti, même si dès la publication du dictionnaire, les termes choisis et les définitions qui y figurent donnent lieu à de nouvelles négociations. Nous pouvons prendre l'exemple du syntagme *développement soutenable* qui est préféré à *développement durable*; ce dernier est la traduction française officielle. Les Verts préfèrent alors parler de *développement soutenable* car il entre en opposition avec *insoutenable*.

Les écologues, les écologistes associatifs ou les écologistes partisans vont ainsi marquer leurs différences scientifiques ou idéologiques, en précisant par différents procédés linguistiques leurs propres opinions ou celles des autres, pour mieux y adhérer ou s'en détacher.

- Que doit-on entendre par «**écologie d'aménagement**» ? Nous désignons par là tout ce qui concerne l'intervention **des spécialistes du milieu de vie** (aspects physiques, chimiques, biologiques et psycho-physiologiques) au regard des diverses pressions que les activités humaines exercent sur ce milieu. (*Aménagement et Nature*, 1974 : 2)

- Par «**effet de la pollution atmosphérique**» nous n'entendons que les effets exprimables en termes économiques, sur l'homme, les animaux, les plantes et les matériaux. (*Aménagement et Nature*, 1973 : 9)

Dans cet exemple nous voyons le besoin éprouvée par la revue *Aménagement et Nature* d'apporter un commentaire métalinguistique à l'expression *effet de la pollution atmosphérique*.

- **Les écologistes**, dans le langage commun, défendent l'essentiel face aux inconscients, bravo, continuez ! (*Combat Nature*, 1985 : 3)

- En dessous de 5 % nous étions marginaux, les grands partis nous ignoraient. À plus de 10 %, **l'écologie (entendez pour eux l'environnement)** est devenue un cheval de bataille des partis. (*Vert Contact*, 1989 : 1)

- La manufacture Michelin a donc acquis, le 1er mars 1990, 400 hectares afin d'y implanter une piste d'essais pour **des pneus dits «verts»**. (note : La raison de cette dénomination est que ces pneus doivent diminuer de 30 % la consommation de carburant, et, partant, de **diminuer d'autant les**

¹⁰FOING, Dominique (dir.) (1994) : *Le livre des Verts, dictionnaire de l'écologie politique*, éd. du Félin, Paris.

émissions de gaz carbonique et autres polluants dans l'atmosphère). (*Combat Nature*, 1993 : 27)

- Notre sincérité politique nous laissait croire que le label «écologiste» était à présent ancré dans l'électorat. La confusion créée par les **pseudo-écologistes** n'est plus à démontrer. (*Vert Contact*, 1993 : 1)

Dans ces énoncés nous pouvons observer la façon dont chaque groupe énonciatif tente de «baliser» le sémantisme de certains termes, en procédant notamment à des mises à l'écart, ou mise à distance de signifiés portés par ces termes, qui sont alors aussitôt récusés et donc attribués aux «autres énonciateurs». Le dernier exemple montre bien comment *Les Verts*, sous-entendu «nous, les vrais écologistes», se distinguent de ce qu'ils nomment des *pseudo-écologistes*.

CONCLUSION

Tout comme les écologistes se sentent dépossédés de leur terminologie les écologistes qui ont adopté une partie de cette terminologie scientifique se sentent dépossédés et trahis de leur idées lorsqu'ils voient la presse généraliste ou la publicité reprendre ou créer des termes de «leur» domaine en y rattachant des signifiés différents (voire opposés). Cela illustre bien le fait que les mots ne sont pas la propriété des domaines dans lesquels ils évoluent. Les mots sont des électrons libres qui se dispersent au gré des changements sociaux et/ou techniques, c'est-à-dire d'éléments extra-linguistiques malgré des instances de normalisation. «L'essentiel des échanges est soumis à un incessant remaniement fait d'actualisation et de réaménagements des catégorisations» (Gaudin 1993 : 196).

Nous avons pu voir que les différents acteurs du domaine étudié déploient des stratégies discursives qui font appel à des besoins lexicologiques et terminologiques spécifiques, dont nous n'avons abordé ici qu'une petite partie.

Chaque groupe énonciatif qu'il cherche à se différencier ou au contraire à se «fondre» dans un discours commun, s'inscrit nécessairement dans un rapport avec «l'autre» et dans une histoire socioculturelle spécifique qu'il nous est nécessaire de prendre en compte.

On pourrait dire en conclusion qu'il existe alors des cycles d'interaction qui se croisent, se succèdent et s'influencent mutuellement créant une circulation de signes : motivation, démotivation, remotivation à laquelle il ne faut pas oublier d'ajouter les processus de normalisation qui participent également au réglage de sens de ces cycles.

RÉFÉRENCES

- DUBOIS, Jean et Louis GUILBERT (1961) : «Formation du système préfixal intensif en français moderne et contemporain», *Le français moderne*, avril 1961, pp. 87-111.
- GAUDIN, François (1993) : *Pour une socioterminologie*, Université de Rouen, 524 p.
- GUESPIN, Louis (1991) : «La circulation terminologique et les rapports science, technique, production», *Terminologie et sociolinguistique*, 18, Université de Rouen, pp. 59-79.
- JACOBI, Daniel (1990) : «Les séries superordonnées dans le discours de vulgarisation scientifique», *Langages*, 98, juin 1990, pp. 103-121.
- JACOBI, Daniel (1986) : *Diffusion et vulgarisation — itinéraires du texte scientifique*, Paris, Les belles lettres, 182 p.

MOTIVATION ET TRANSPARENCE DES EMPRUNTS GRÉCO-LATINS EN FRANÇAIS ET EN ALLEMAND

Christine JACQUET-PFAU et Marie-Anne MOREAUX

CERTAL et Collège de France et CERTAL et INALCO, Paris, France

L'un des thèmes de recherche du CERTAL (Centre d'Études et de Recherche en grammaire et Traitement Automatique des Langues) est l'étude des emprunts dans une perspective de traitement automatique. Les recherches de cette équipe implantée à l'INALCO (Institut National des Langues et Civilisations Orientales) se situent dans un environnement fortement multilingue et mettent l'accent sur l'étude du système morphologique des langues traitées : tchèque, arabe, français et allemand pour les membres de l'équipe, anglais, italien, polonais, berbère et malais dans le cadre de coopérations et de thèses. Elles ont pour but de développer des applications en indexation automatique et des outils terminologiques.

Envisagé dans un sens large, le niveau morphologique englobe tous les systèmes participant à la construction des unités lexicales apparaissant dans un texte, c'est-à-dire non seulement le système de flexion et de dérivation, mais également tout système de formation des lexèmes. L'étude fine des unités lexicales montre que leur forme et leur structure offrent souvent d'immenses possibilités d'auto-détermination de valeurs morpho-syntaxiques et morpho-sémantiques. L'équipe considère ainsi :

- que l'étape morphologique est une étape cruciale dans tout système d'analyse automatique car, bien conduite, elle permet de réduire la complexité des étapes suivantes;
- que cette étape doit reposer prioritairement sur l'application d'ensemble de règles et de techniques de reconnaissance de formes décrivant la structure et les propriétés des unités lexicales plutôt que sur des dictionnaires qui, par définition, ne peuvent jamais être complets.

Le développement d'un analyseur morphologique ayant recours à ce type de techniques nécessite de partitionner théoriquement le lexique en deux ensembles : les termes autochtones (p.ex. romans pour le français, germaniques pour l'allemand) et les termes non autochtones, qui peuvent à leur tour être partitionnés en différents sous-systèmes : emprunts aux langues vivantes, formations gréco-latines. Cette partition minimale présente plusieurs avantages :

- elle concourt à augmenter l'efficacité de l'analyse morphologique en permettant le traitement des interférences de formes appartenant aux différents systèmes par des

techniques relativement simples (reconnaissance de formes, schémas graphémiques et syllabiques);

- elle permet de faire des hypothèses, à partir de ses constituants, sur une unité lexicale qui ne peut être reliée à aucune forme lexicale attestée de la langue traitée. Au cas où aucune hypothèse ne peut être faite, le système est au moins informé de la raison de l'échec s'il est capable de reconnaître que l'unité n'appartient pas à l'ensemble autochtone;
- les unités lexicales constituées de formants gréco-latins sont la plupart du temps des lexèmes nominaux ou adjectivaux et dans les textes techniques bien souvent des unités terminologiques. Leur reconnaissance est donc d'une importance capitale pour la reconnaissance et l'extraction de termes-clés.

L'un des corpus utilisés par le CERTAL pour vérifier ses hypothèses est constitué de textes relevant du domaine de l'écologie, domaine choisi en raison de sa pluridisciplinarité : il intéresse aussi bien la biologie, la géographie, la géologie, que la chimie, la physique, l'économie... Appartenant à ce corpus, l'*Atlas zur Ökologie* de Dieter Heinrich et Manfred Herget, et sa traduction par Josselin Mercier, l'*Atlas de l'écologie*, nous ont permis de tester l'analyse que nous présentons des emprunts gréco-latins en français et en allemand, et que l'on pourrait étendre à d'autres langues. Cet ouvrage scientifique mais dont l'objectif didactique (aidé par la mise en pages, l'alternance page d'illustrations/page de texte, le strict parallélisme entre le texte original et la traduction...) permet en effet de repérer aisément la nomenclature de base du domaine et d'offrir, dans les deux langues, un échantillonnage important de formations gréco-latines.

MOTIVATION ET TRANSPARENCE

Les termes «motivation» et «transparence» sont la plupart du temps utilisés comme termes synonymes, «motivation» l'emportant dans l'usage. La notion ainsi désignée, opposée à celle d'arbitraire, reste souvent floue et l'on ne discerne pas toujours clairement s'il s'agit de rendre compte de relations entre unités de nature purement linguistique ou du lien qui existe entre une entité de la réalité et sa dénomination.

Nous allons préciser ce que nous entendons par «motivation» et «transparence», que nous n'employons pas de manière synonyme. Notre approche vise à établir des critères permettant de développer, à l'usage d'une machine, une description de l'organisation du lexique qui prenne en considération certains phénomènes de créativité lexicale. Cette description se situe dans un premier temps sur un plan strictement linguistique et cherche à rendre compte du fait qu'il est possible de «calculer» les propriétés grammaticales et lexicales d'un grand nombre de formations complexes à partir de leurs éléments, autochtones ou non. Cela suppose que ces éléments puissent être identifiés, que les modèles de construction des unités complexes soient connus, mais également, même si l'on admet que l'interprétation d'une unité lexicale dépend de son contexte, qu'il existe un minimum de signification lexicale conventionnelle, parfois très générale, relativement stable sans laquelle aucune interprétation (humaine ni automatique) n'est possible. Ce n'est que dans un second temps que les unités complexes permettent d'orienter vers l'entité qu'elles dénomment, et cela d'autant mieux lorsqu'elles sont descriptives.

Ainsi, même s'il est banal de constater que toutes les unités lexicales présentant des analogies sur le plan de la forme n'en présentent pas sur le plan du contenu et que toute

unité lexicale ne peut être interprétée à partir de ses constituants, il faut cependant également reconnaître que la forme, dans une certaine mesure, contraint la signification et qu'inversement la signification contraint la forme. La grande majorité des néologismes, par exemple, est construite à partir d'éléments préexistants (dans la langue ou dans une autre langue) selon des schémas connus, ce qui est une condition absolument nécessaire à leur décodage. Aussi ferons-nous la distinction entre :

◆ **motivation** (vs démotivation), que nous restreignons à la motivation morphématique. Elle repose uniquement sur des critères de forme (graphie, forme des morphèmes) et permet de mettre en relation le signifiant d'une unité complexe avec celui d'unités moins complexes (famille de mots, mais pas obligatoirement famille de sens). Une unité lexicale peut être :

- totalement motivée : en fr. *préavis, clairvoyance*; en all. *Vorfeld, Sprachwissenschaft*;
- partiellement motivée : en fr. *préface, cécité*; en all. *Himbeere*.

Qu'une unité lexicale soit motivée ne signifie pas nécessairement que sa signification puisse être construite à partir de celle de ces constituants : elle peut être lexicalisée, certains des constituants peuvent avoir perdu leur autonomie sémantique.

◆ **transparence** (vs opacité), qui est la propriété d'une unité lexicale à laisser «transparente» sa signification à travers sa forme et donc à pouvoir être interprétée de manière compositionnelle. La condition, nécessaire mais non suffisante, de la transparence est la stabilité des signifiants. Pour être transparente une unité lexicale doit être motivée, mais inversement, une unité motivée peut n'être que partiellement transparente ou être totalement opaque :

- transparence totale : en fr. *prédécoupage*; en all. *Vorarbeit*;
- transparence partielle : en fr. *préface*; en all. *Vorhang*;
- opacité : en fr. *président*; en all. *aufhören*.

Dans cette perspective, toute réforme orthographique devrait être menée avec extrêmement de précaution et après évaluation de toutes les conséquences qu'elle peut entraîner; p.ex. la modification de la forme habituelle d'unités lexicales risque de conduire à leur démotivation et donc à leur opacification. Ainsi l'orthographe de nombreux mots français qui paraît a priori aberrante se justifie parfaitement par ces notions de motivation et de transparence.

Alors que les signifiants des constituants ont tendance à fusionner à l'intérieur des unités lexicales complexes du français, ils restent remarquablement stables dans celles de l'allemand et l'on peut dire, selon la formule de J.-M. Zemb que «l'allemand est une langue indéradicable». L'une des causes, sinon la cause principale, de cette différence est à chercher dans l'ordre d'association que l'allemand et le français utilisent pour organiser les éléments de la phrase d'une part et les éléments lexicaux en lexèmes d'autre part.

En allemand, l'ordre d'assemblage déterminant/déterminé que Tesnière nomme centripète se retrouve à tous les niveaux d'organisation des éléments de la phrase, du syntagme aux unités lexicales. Le français est par contre fondamentalement centrifuge

ESSAI DE TYPOLOGIE DES EMPRUNTS GRÉCO-LATINS EN ALLEMAND ET EN FRANÇAIS

L'utilisation par une langue d'un signe appartenant à une autre langue (un morphème, un lexème, une construction syntaxique) constitue un emprunt. Ce phénomène est connu par toutes les langues à partir du moment où elles ont été au contact avec au moins une autre langue, et ce tout au cours de leur histoire : en ce sens l'emprunt représente une sorte de mémoire transversale, même si cette dernière est plus ou moins consciente et transparente pour les locuteurs, en fonction de plusieurs facteurs parmi lesquels nous retiendrons essentiellement la nature de l'emprunt, sa structure et son origine.

Les emprunts aux langues mortes (grec et latin) constituent, non seulement pour les langues romanes mais également pour des langues de groupes différents comme le français, l'allemand, l'anglais ou encore le tchèque, un véritable fonds commun dont il conviendrait, au-delà des considérations étymologiques traditionnelles, de tirer le plus grand parti. Ce système non autochtone forme un lexique commun, entre autres, au français et à l'allemand, que J.-M. Zemb appelle *koiné*.

Si l'emprunt est une notion couramment évoquée, sa définition n'est pas aussi «transparente» qu'il y paraît au premier abord. Le champ qu'elle recouvre est en effet si vaste et si diversifié qu'il est nécessaire d'en préciser les acceptions, en général, et l'acception plus particulière pour le sujet qui nous intéresse ici.

Il est possible d'établir, dans les deux langues, une distinction historique entre *mots hérités* et *emprunts* (ces derniers correspondant à ce que la tradition appelle formation savante) :

- les *mots hérités* correspondent au passage inconscient, par voie orale et donc selon un processus plus ou moins complexe, d'un lexème — d'un signifié et d'un signifiant — d'une langue à une autre. Ces mots sont soumis aux transformations «naturelles» et progressives de la langue. En français, les mots hérités sont en général de datation ancienne et sont issus du latin. Ils ne sont plus motivés en synchronie sous leur forme dite «populaire» (héritée) mais le sont sous leur forme dérivée dite «savante», recrée plus tard directement sur la base latine, cette seconde forme restant stable :

- *hôtel* est le résultat de modifications successives du bas-latin *hospitale* (*ostel* → *hospel* → *hôtel*) (forme savante : *hospitalier*)
- l'évolution phonétique normale du latin *vitrum* a donné *voirre* (1155) puis *voarre* (1380) pour aboutir à *verre* (forme savante : *vitre*)
- germ. **kuningaz* > vha. *kuning* > mha. *künic*, *künc* > nha. *König*

- les *emprunts* gréco-latins correspondent à un processus conscient, mis en oeuvre dès qu'il a été nécessaire de créer des termes pour désigner une nomenclature spécialisée. L'emprunt, à la différence de l'héritage, plus ancien, est fait au grec ou au latin classique. Relevant d'une langue écrite et donc normée, il présente une plus grande stabilité phonographémique. Il consiste à utiliser, parmi le stock lexical du grec ou du latin, soit des lexèmes :

gr. *bibliothêkê* → lat. *bibliotheca* → fr. *bibliothèque* → all. *Bibliothek*
gr. *pinakothêkê* → lat. *pinacotheca* → fr. *pinacothèque* → all. *Pinakothek*

soit des formants :

fr. *insectivore* et all. *insectivor* d'après le latin *carnivorus*
géophysique (fin XIX^e) emprunté à l'anglais *geophysics* (1889)
géopolitique (1924) emprunté au suédois *geopolitisk* (1889), peut-être par l'anglais
geopolitics (1904),
géochimie emprunté à l'allemand *Geochemie* (1838),
le formant *geô-* ayant été morphématisé d'après le grec *geôgraphia* (géographie)

Nous voyons par ailleurs à travers ces quelques exemples que l'histoire des mots empruntés ou construits à partir de formants grecs et latins est souvent complexe et qu'il n'est pas toujours facile d'en connaître le parcours, le mot transitant par une, voire plusieurs langues, avant d'être emprunté. Mais la dernière série montre également que ces termes ont pour caractéristique commune d'avoir la même origine et d'être construits d'après le même modèle. Ainsi, au cours des siècles, a été créée une zone lexicale internationale, où les termes n'appartiennent pas plus à une langue qu'à une autre, à moins qu'ils n'appartiennent qu'à une seule langue, si l'on veut bien considérer que les différents domaines scientifiques se sont peu à peu créés chacun une langue, motivée et transparente :

all. <i>Linguistik</i>	→	fr. <i>linguistique</i>
all. <i>Ökologie</i>	→	fr. <i>écologie</i>
all. <i>Makrobiotik</i>	→	fr. <i>macrobiotique</i>
fr. <i>radioactivité</i>	→	all. <i>Radioaktivität</i>

La volonté et le besoin de créer des vocables non ambigus et compréhensibles par le plus grand nombre de langues ne datent pas d'aujourd'hui. Dès le XIV^e s., et jusqu'au XVI^e s., les savants entreprennent une véritable relatinisation de la langue française, doublant dans un même temps le volume du vocabulaire. La redécouverte du grec fournit de nombreux mots et, à partir du XVIII^e s., les formes gréco-latines se généralisent dans le langage scientifique (celui de la botanique avec Linné et Jussieu, de la chimie avec Guyton de Morveau et Lavoisier). Ce phénomène connaît sa pleine ampleur au XIX^e s. avec le développement des sciences et de l'industrie.

Les formations gréco-latines présentent certaines propriétés :

- formellement, elles sont motivées;
- sémantiquement, elles sont plus ou moins transparentes;
- fonctionnellement, elles sont utilisées ou utilisables par un large groupe de langues occidentales.

Ce vocabulaire dit «savant» est véhiculé avant tout par la langue écrite et assure une fonction précise. L'une de ses principales caractéristiques est donc sa stabilité. Sa graphie

n'a été soumise qu'à de faibles modifications, et ce n'est que depuis une époque récente que les contraintes pesant sur la nature des formants (homogénéité) et leur nombre dans un même lexème (la référence au modèle limitait la formation à deux éléments lexicaux) commencent à évoluer. La graphie des unités répond, à l'intérieur de chaque langue, à quelques règles de transcription phono-graphémique et ne subit pas les modifications du système général. D'une langue à l'autre on peut retrouver des règles de transcription. En ce sens on peut considérer que les emprunts gréco-latins constituent à la fois un sous-code de la langue et un code «international», relativement facile à reconnaître (*Tableau 1*).

grec / transcription	latin	allemand	français	exemple
[-]F-/ ph-	[-]ph-	[-]ph-	[-]ph-	phull- phyllanthes [Chloro]phyll [chloro]phyll
k- / k-	c-	k-	c-	klima clima Klima climat
[-]q- / th-	[-]th-	[-]th-	[-]th-	bibliothêkê bibliotheca Bibliothek bibliothèque
-ia / -ia	-ia	-ie	-ie	geôgraphia geographia Geographie géographie
	-ci-	-zi-	-ci-	-cide < caedere Insectizid insecticide
-iko"/ -ikos	-icus	-ik /-isch	-ique	-ikos -icus -isch -ique
	[-]vor-	[-]vor-	[-]vor-	vor[are] Insektivore insectivore

Tableau 1 : Quelques correspondances phono-graphémiques

Il y a donc tout intérêt à maintenir cette stabilité graphique, l'ensemble des formants gréco-latins et leur schéma de construction permettant de construire des termes en général transparents et non ambigus pour les spécialistes du domaine et faciles à mémoriser. On sait bien que les spécialistes d'un domaine, même s'ils sont locuteurs de la même langue et même s'il existe des dénominations autochtones des concepts de leur domaine de spécialité, entre eux, utilisent de préférence les dénominations d'emprunt. À plus forte raison s'ils sont de langues différentes. Ainsi peuvent-ils avoir non seulement des concepts

communs, mais des dénominations communes de ces concepts. Il existe en effet un lexique, une terminologie «internationalisable» (Cottez 1989 : XXII).

On a utilisé des formants grecs ou latins pour créer, à partir d'un mot déjà emprunté et faisant alors office de modèle. La motivation peut être forte chez le lexicographe :

- azote est un terme créé par les chimistes Guyton de Morveau et Lavoisier en 1787, «à partir de l'*a-* privatif du grec et de *zôê* (vie)», d'après Morveau lui-même; ce terme s'est substitué à *nitrogène* (l'anglais, lui, a conservé le terme *nitrogen*) pour désigner un gaz dont la principale caractéristique est d'être un gaz impropre à la respiration ou un gaz dans lequel la vie ne peut se développer (all. *Azot / Azote* [Stickstoff]).

Mais, dans la plupart des formations, et de plus en plus au fil du temps du fait de l'augmentation du stock lexical terminologique de la langue qui emprunte, «la création s'opère dès lors, non plus par référence à la motivation étymologique, mais par référence à un modèle fonctionnel spécifique dans un champ sémantique particulier» (Guilbert 1975 : 231). La connaissance du paradigme s'avère alors essentielle pour déchiffrer le code :

- sur le modèle de *bibliothèque* (emprunté au latin 1493) ont ainsi été créés des composés savants tels que *discothèque* (1928), *ludothèque* (1970), *vidéothèque* (1970), *modothèque*, *audiothèque*, *pochothèque*, *inathèque* (néologismes).

Une dernière remarque s'impose quant à l'origine des formants. Il semble que la terminologie ait toujours préféré remonter à la source même de son modèle, à savoir le grec. L'analyse du corpus constitué par les termes-clés indexés à la fin de l'*Atlas de l'écologie* confirme que, dans ce domaine de spécialité, les constructions grecques sont effectivement plus nombreuses que les constructions latines, ces dernières étant formellement moins marquées (moins motivées) terminologiquement. Ainsi le formant *hydr-* contribue à former sept termes tandis que le formant *aqua-* n'apparaît que dans un seul terme (Tableau 2). Quant au formant latin qui correspond au grec *poly-* et que nous trouvons normalement sous deux formes, *multi-* et *pluri-*, il n'apparaît dans aucune formation de ce corpus (Tableau 3).

formant latin	formant grec
aquatique / aquatisch	hydraulique / Hydraulik
	hydro-biosphère / Hydro-Biosphäre
	hydrocarbure / Kohlenwasserstoff
	hydrolyse / Hydrolyse
	hydrosphère / Hydrosphäre
	hydrophytes / Hydrophyten
	hydrosystème / Hydrosystem

Tableau 2 : Les formants *aqua-* et *hydro-* dans l'index de l'*Atlas de l'écologie / Atlas zur Ökologie*

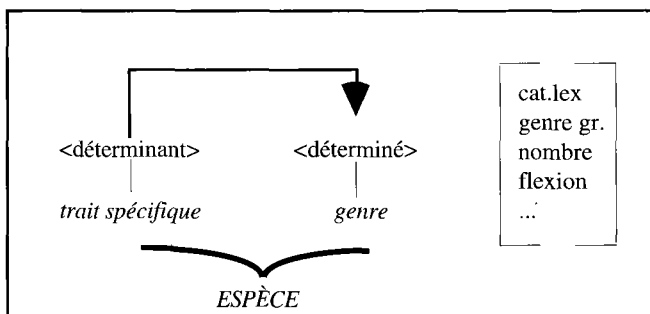
formant latin	formant grec
	polymorphisme / Polymorphismus
	polychlorobuphényles (PCB) / polychlorierte Biphenyle (PCB)
	polyplôidie / Polyploidie
	polysaprobe / polysaprob

Tableau 3 : Les formants *multi-* / *pluri-* et *poly-* dans l'index de l'Atlas de l'écologie / Atlas zur Ökologie

Certains auteurs ont été jusqu'à «se poser la question de savoir pour quelle raison le Prob [Petit Robert] comme les autres dictionnaires de langue préfère l'indication «mot emprunté au latin» à «internationalisme (d'origine inconnue) formé sur la base du fonds gréco-latin» ou — ce qui serait plus précis mais très difficile à décider dans la plupart des cas — «mot basé sur le fonds gréco-latin, adopté par l'intermédiaire de l'anglais, de l'allemand ou d'une autre langue» (Schmitt 1991 : 124). Nous gardant bien d'établir cette confusion entre étymologie et analyse synchronique, nous allons montrer que l'utilisation de formants gréco-latins permet de motiver et de rendre transparente la terminologie d'un domaine, qui doit être univoque et, plus que jamais, internationale.

STRUCTURE INTERNE ET INTERPRÉTATION DES FORMATIONS GRÉCO-LATINES

La grande majorité des unités lexicales constituées de formants gréco-latins doit être interprétée, dans les deux langues, selon le modèle des constructions déterminatives. La relation qui existe entre les composants est une relation de détermination au sens où l'entité dénommée par le nouveau lexème est toujours une espèce de l'entité dénommée par le *déterminé*, le *déterminant* apportant le trait spécifique. Ces unités lexicales ont ainsi des propriétés descriptives et classificatoires. Leur structure est toujours binaire et, quelques cas mis à part, centripète, donc inverse de la structure générale du français. Les propriétés grammaticales et syntaxiques de l'unité lexicale (catégorie lexicale, genre grammatical, catégorie flexionnelle) sont normalement apportées par le déterminé. La structure générale des unités déterminatives peut être schématisée de la manière suivante :



- Le déterminé peut être un lexème, c'est-à-dire un élément qui non seulement possède une autonomie sémantique, ce qui est le cas de la plupart des formants gréco-latins, mais également une autonomie syntaxique :

<i>paléogéographie</i>	<i>Paläogeographie</i>
<i>biogéographie</i>	<i>Biogeographie</i>
<i>hydrogéographie</i>	<i>Hydrogeographie</i>
<i>zoogéographie</i>	<i>Zoogeographie</i>
<i>phytogéographie</i>	<i>Phytogeographie</i>

Le déterminé indiquerait ainsi que la branche scientifique désignée doit être, à tort ou à raison, considérée comme un domaine de la géographie, ce qu'étudie ce domaine étant précisé par le déterminant.

- Le déterminé est un élément lexical n'ayant pas de forme syntaxique autonome :

acaricide	akarizid	Akarizid
insecticide	insektizid	Insektizid
herbicide	herbizid	Herbizid
fongicide	fungizid	Fungizid
pesticide	pestizid	Pestizid
biocide	biozid	Biozid

L'élément *-cide / -zid* a pour contenu sémantique [DESTRUCTION]. L'unité globale désigne un moyen de destruction (nom d'agent) ou signifie «relatif à un moyen de destruction» (adjectif relationnel), le déterminant représentant l'objet de la destruction. La distinction nom d'agent/adjectif relationnel n'est pas marquée morphologiquement en français, mais est perceptible en allemand grâce à la majuscule et à la flexion.

Dire que la structure est binaire ne signifie pas que ces unités lexicales sont nécessairement formées de deux éléments lexicaux car chacun d'entre eux peut lui-même avoir une structure déterminative et être constitué d'éléments ayant une structure déterminative. Si l'on représente les composants par les lettres de l'alphabet *a, b, c, ...*, la hiérarchie des relations de détermination qui existe entre les composants peut être représentée par des expressions parenthésées :

a (b c) : (*anéo(dynamo(mètre))*) / (*Anemo(dynamo(meter))*), qui est une sorte de *dynamomètre*, le terme *dynamomètre* étant lui-même de type déterminatif.

(a b) c : ((*phyto(plancto))toxine*) / ((*Phyto(plankton))toxine*), qu'il faut interpréter comme désignant un type de toxines dû au plancton végétal et non un type de toxine dû au plancton et s'attaquant aux végétaux.

(a b) c : ((*ana(biot))ique*) / ((*ana(biot))isch*); ((*anti(biot))ique*) / ((*Anti(biot))ikum*), ((*anti(biot))isch*). Dans ce cas le déterminé est un suffixe.

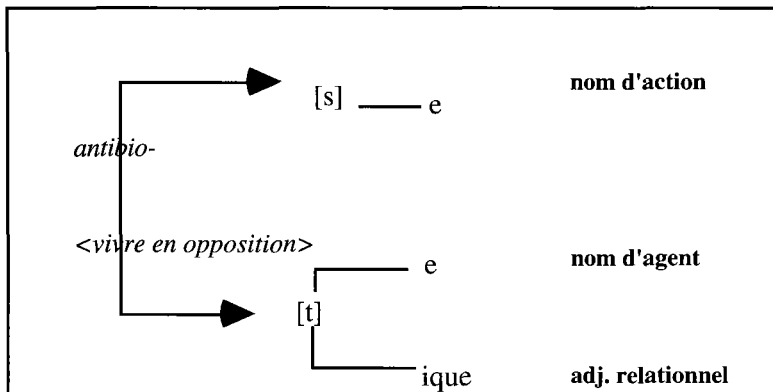
L'identification de la structure interne et de la hiérarchie des relations entre les constituants est indispensable à l'interprétation de ces unités, puisque leur signification n'est pas simplement la somme des contenus de leurs constituants. Ainsi, si en méconnaissant la

structure du terme *antibiotique*, on le décompose en deux constituants *anti* et *biotique* et qu'on l'interprète à partir de leur signification sans tenir compte du mode de relation qui les lie, on risque d'aboutir à la signification suivante «contre ce qui est relatif aux organismes vivants».

L'interprétation est particulièrement délicate lorsqu'il existe dans le lexique une unité lexicale autonome dérivée dont la forme peut être reconnue comme dernier constituant. Il s'agit alors de déterminer si l'unité est un composé dont le dernier constituant est un dérivé ou si elle est ce que nous appellerons un *dérivé à base composée*. Dans le premier cas la hiérarchie des relations est du type a(bc), dans le second du type (ab)c. Le déterminé est alors un suffixe qui possède essentiellement une fonction de catégorisateur et dont la signification est extrêmement générale.

Pour être interprété comme «relatif à l'antibiotique, c'est-à-dire celui dont le mode de vie est l'antibiose», le terme *antibiotique* doit être replacé à l'intérieur d'une série doublement structurée :

- syntagmatiquement : *antibiose*, *antibiote*, *antibiotique* sont respectivement le nom d'action, le nom d'agent et l'adjectif relationnel construit sur la base *antibio-* :



- paradigmatiquement : comme dans *anabio-se /-te /-tique*, *probio-se /-te /-tique*, *métabio-se /-te /-tique*, *ymbio-se /-te /-tique*, *bio*, le déterminé de la base, a pour contenu sémantique [VIE, VIVRE], le déterminant spécifiant le genre de vie, la manière de vivre :

(X(bio))y	
bio	[VIE/VIVRE/VIVANT]
X	spécificateur
y	catégorisateur

L'appartenance des formations gréco-latines à des séries dérivationnelles est fréquente et peut être justifiée étymologiquement. Si l'on examine, à partir des termes *géographie*, *géographe*, *géographique*, all. *Geographie*, *Geographe*, *geographisch*, la longue liste des formations en *-graphie*, *-graph(e)*, on constate le même phénomène, les unes correspondant au nom d'action, les autres au nom d'agent. La forme du nom d'action pouvant également être utilisée pour nommer le résultat, *-graphie* entre alors en concurrence avec *-gramm(e)*.

Au-delà de la mauvaise interprétation d'une unité particulière, la non prise en compte de ces séries, due au fait qu'elles ne sont plus perceptibles au travers de suffixes spécifiques, risque de provoquer de fausses morphématisations conduisant à l'opacification de séries d'unités et à la construction de nouvelles unités non transparentes. Cette opacification peut n'être que momentanée, si le nouvel élément lexical séparé de la base qui a permis de le créer prend un contenu différent qui se régularise dès qu'il est utilisé pour construire de nouvelles séries de composés. On peut actuellement discerner un formant *-graphie* ayant le contenu [MÉTHODE, PROCÉDÉ] même s'il est encore possible d'y discerner un archilèxème <ACTION DE TRACER>.

La relation de détermination est une relation très générale qui peut s'établir selon différents modes. Puisque toute marque de relation explicite est effacée dans la concaténation des constituants, on ne sait pas a priori quel est le type de la relation qui unit le déterminant et le déterminé. Le trait spécifique peut correspondre à une propriété caractérisant la forme ou la fonction de l'entité que l'on veut dénommer : p.ex. relation d'attribution dans *macrophage*, relation 'objet, patient / prédicat' dans *ombrophil(e)*, relation 'prédicat / sujet, acteur' dans *Phagozyt*. Il serait donc nécessaire pour chacun de ces formants néo-classiques de déterminer quelle catégorie de déterminant il peut accepter et inversement de quelles catégories d'unités il peut être le déterminant afin d'établir des classes d'éléments selon les relations qu'ils sont susceptibles d'entretenir.

En allemand *Geographie* possède un doublet autochtone, sorte de calque, qui est donc lui-même un composé déterminatif, *Erdkunde*. Les deux termes sont donnés dans tous les dictionnaires comme équivalents. On remarque cependant que dans la série citée ci-dessus *Erdkunde* ne peut remplacer *Geographie* que dans le terme *Paläogeographie*. Dans les autres cas, lorsqu'il existe un doublet, celui-ci est hybride : *Tiergeographie* pour *Zoogeographie*, *Pflanzengeographie* pour *Phytogeographie*, *Erdkunde* étant trop motivé et trop transparent pour avoir un contenu se rapportant à tous les phénomènes et les êtres en relation avec la surface du globe.

Il semble ainsi que les éléments gréco-latins, qu'ils soient simples ou complexes, sont avant tout utilisés pour leur contenu lexical qui, bien que fort, reste très général. C'est probablement ce qui fait leur souplesse d'utilisation, donc leur grande productivité et leur capacité à être employés dans tous les domaines de la langue de la connaissance.

RÉFÉRENCES

- BAILLY, A. (1950) : *Dictionnaire grec-français*, Paris, Hachette, 2230 p.
- BERGMANN, Rolf (1995) : «'Europäismus' und 'Internationalismus'. Zur lexikologischen Terminologie», *Sprachwissenschaft* 20/3, Heidelberg, Universitätsverlag C. Winter, pp. 239-277.
- BRAUN, Peter (Hrsg) (1979) : *Fremdwort-Diskussion*, coll. «UTB» 797, München, Wilhelm Fink Verlag, 363 p.
- COTTEZ, Henri (1989) : *Dictionnaire des structures du vocabulaire savant. Éléments de formation*, 4^e éd., coll. «Les usuels du Robert», Paris, Le Robert, 515 p.
- FLINCK, Hansi-R. (1996) : *Fachsprachen : Einführung und Bibliographie*, coll. «UTB» 483, Tübingen und Basel, A. Francke Verlag, 361 p.
- GAFFIOT, Félix (1934) : *Dictionnaire latin-français*, Paris, Hachette, 1719 p.
- GUILBERT, Louis (1975) : *La créativité lexicale*, coll. «Langue et langage», Paris, Larousse, 285 p.
- GUIRAUD, Pierre (1978) : *Les mots savants*, 1^{re} éd. 1968, coll. «Que sais-je ?», n° 1325, Paris, PUF, 128 p.
- HEINRICH, Dieter et Manfred HERGT (1990) : *Atlas zur Ökologie*, München, Deutscher Taschenbuch Verlag, 286 p.
- HEINRICH, Dieter et Manfred HERGT (1993) : *Atlas de l'écologie*, coll. «La Pochothèque», Paris, Librairie Générale Française, 286 p.
- HOPPE et al. (1987) : *Deutsche Lehnwortbildung*, Tübingen, Forschungsberichte des Institut für deutsche Sprache, n° 64, 467 p.
- MOREAUX, Marie-Anne (1994) : «Détection, segmentation et interprétation des noms multilexicaux allemands», P. Bouillon et D. Estival (dir.), *Actes du Colloque «Journées Noms composés, Aspects multilingues de la composition nominale»*, Genève, pp. 88-102.
- PICOCHÉ, Jacqueline (1991) : *Dictionnaire étymologique du français*, coll. «Les usuels du Robert», Paris, Le Robert, 827 p.
- REINHARDT, W., KÖHLER, C. et G. NEUBERT (1992) : *Deutsche Fachsprache und Technik*, Band 3, Hildesheim / Zürich / New York, Olms, 174 p.
- REY, Alain (dir) (1992) : *Dictionnaire historique de la langue française*, 2 tomes, Paris, Le Robert, 2383 p.
- SCHAEFER, Matthias (1992) : *Ökologie (Wörterbuch der Biologie)*, coll. «UTB» 430, Iena, Fischer Verlag, 433 p.

- SCHMITT, Christian (1991) : «L'Europe et l'évolution des langues de spécialité», *Terminologie et traduction*, n° 2, Commission des communautés européennes, pp. 115-127.
- WAHRIG-BURFEIND, Renate (1996) : *Fremdwörter-Lexikon*, Gütersloh, Bertelsmann Lexikon-Verlag, CD-Rom.
- ZEMB, Jean-Marie (1978) : *Vergleichende Grammatik Französisch-Deutsch*, Teil I. *Comparaison de deux systèmes*, Mannheim / Wien / Zürich, Bibliographisches Institut, Duden & Paris, Didier, 897 p.
- ZEMB, Jean-Marie (1984) : *Vergleichende Grammatik Französisch-Deutsch*, Teil II. *L'Économie de la langue et le Jeu de la parole*, Mannheim / Wien / Zürich, Bibliographisches Institut, Duden & Paris, Didier, 975 p.
- ZEMB, Jean-Marie (1997) : «La coexistence proléifique de deux systèmes lexicaux en allemand», Georges Kleiber et Martin Riegel (dir), *Les Formes du sens : Études de linguistique française, médiévale et générale offertes à Robert Martin à l'occasion de ses 60 ans*, coll. «Champs linguistiques. Recueils», Paris, Duculot, pp. 423-436.

LA DÉRIVATION EN MÉMOIRE

Daniel BLAMPAIN

ISTI - TERMISTI, Bruxelles, Belgique

Partant du constat trivial que le locuteur a *la mémoire des mots*, autrement dit qu'il dispose d'une compétence lexicale qui lui est propre, on peut se poser, à l'examen de ses performances, des questions qui engagent une réflexion sur la conception de la langue, de sa pratique et de son enseignement et qui pourraient contribuer à redynamiser le français à l'échelle internationale, notamment en remédiant à cette timidité néologique, que les allophones perçoivent bien et que M. Picone (Université d'Alabama) vient encore de caractériser dans son ouvrage *Anglicisms, Neologisms and Dynamic French* (1996).

Nous poserons le problème en limitant la compétence lexicale à son application au lexique examiné en tant qu'objet de dérivation. Ceci a une double implication :

1. Sont utilisés les concepts chomskyens de *compétence grammaticale*, qui correspond à la capacité de tout individu de produire des phrases (créativité) et de porter des jugements intuitifs sur les phrases, et de *performance*, qui s'applique au résultat de la mise en oeuvre de la compétence.

Les lexicologues ont annexé avec profit ces concepts. D. Corbin (1987 : 46-54) a bien montré la productivité du concept, notamment en le restreignant à la *compétence dérivationnelle*, c'est-à-dire à la compétence applicable aux mots construits «dont le sens prédictible est entièrement compositionnel par rapport à la structure interne, et qui relève[nt] de l'application à une catégorie lexicale majeure [base] d'une opération dérivationnelle [...] associant des opérations catégorielles sémantico-syntaxique et morphologique» (1987 : 6).

2. Nous aurons surtout recours au concept linguistique et pragmatique d'*acceptabilité*, qu'il conviendrait d'approfondir. Si l'on peut admettre que la dérivation repose sur un certain nombre de régularités identifiables, la *lexicalité* d'un mot est à déterminer selon des paramètres intrinsèques et extrinsèques plus complexes que la *grammaticalité* d'une phrase et les seuils d'acceptabilité sont à examiner de plus près. C'est ici que se pose le problème de la *dérivation en mémoire*.

L'ACCEPTABILITÉ D'UN MOT CONSTRUIT PAR DÉRIVATION

1. Le mot, objet du jugement d'acceptabilité, est à distinguer de la phrase. Soit la phrase : *Le colloque que le réseau LTT, que connaissent bien les universités*

francophones, organise à Tunis, est un succès. Cette phrase est grammaticale, mais inacceptable. Le jugement est ici porté sur la combinatoire que la phrase actualise par rapport à l'accessibilité de son sens dans des conditions normales d'échange.

Dans le cas du mot construit, le jugement portera sur une unité plus restreinte que la phrase, davantage exposée à l'irrégularité que la syntaxe et non soumise à des faits de syntaxe (critère de distinction des morphèmes flexionnels et des morphèmes dérivationnels), même si cette unité influe sur la syntaxe, étant donné la puissance de changement de catégorie que possède la dérivation.

Certes, en langue générale, le jugement porte sur l'accessibilité au sens, mais en même temps sera prise en considération l'existence du mot ou de la construction morphématique, si l'on prend comme unité d'examen le morphème qui, rappelons-le, n'a pas l'autonomie syntaxique du mot.

Soit la phrase suivante, prononcée par un commissaire européen italien bien connu pour ses interventions médiatiques : *Il y a eu beaucoup d'accrochements entre les pêcheurs français et espagnols* (juillet 1997). Constat pragmatique : nous sommes dans une situation de discours oral, et le locuteur n'a pas conscience de la nouveauté du mot puisqu'il le répète et le récepteur, lui, a éventuellement son attention attirée par *accrochement*, dont le sens est accessible. Contrairement à une hypothèse émise en son temps par Guilbert (1975 : 30), la mise en oeuvre de règles n'est pas plus consciente, en langue générale, dans le cas du mot construit que dans le cas de la phrase.

On n'est pas conscient, autre exemple, d'une règle aussi fondamentale que celle-ci : le sens négatif d'un mot construit ne peut venir que du préfixe. Le locuteur peut difficilement rendre compte de sa compétence et de sa performance. Tel était déjà le point de vue de Chomsky.

Ce qui, par contre, devrait être pris en considération, c'est la particularité des facteurs conditionnant la performance et informant un autre type de jugement d'acceptabilité.

2. La *lexicalité* du mot *accrochement* est perçue immédiatement : les composantes phonétique, dérivationnelle et sémantique sont satisfaisantes. La construction de noms d'action sur base verbale (opération catégorielle par excellence) : $V \rightarrow V (-ment) \rightarrow N$, est ici actualisée. Cette règle, de même que $V \rightarrow V (-age) \rightarrow N$, appartient à un niveau de dérivation régulier, c'est-à-dire qu'elle a le pouvoir de générer une infinité de mots aux propriétés attendues. Nous sommes ici en situation de concurrence affixale ou de prédictibilité flottante. Et ce d'autant plus que les suffixes *-age* et *-ment* n'ont pas de sens propre. Le sens est issu du résultat de l'affixation. *Accrochage* et *décrochage*, *alunissage*, *amarrage*, mais *abaissement*, *acclimatement*, *arraisonnement* (et non *arraisonnage*), *attaque* (et non *attaquement*), mais également *arrachage* et *arrachement*, *atterrissage* et *atterrissement*, *blanchissage* et *blanchissement*, *déchiffrage* et *déchiffrement*, *raccrochage* et *raccrochement*.

La lexicalité eût été remise en question par *accrochage**. La confusion entre *accrochement* et *accroc* aurait posé un problème sémantique.

3. Le choix de *accrochement* se réfère à un fait de mémoire, non seulement sémantique, mais statistique. La règle $V \rightarrow N$ est la plus fréquente. La fréquence des noms d'action en *-age* et *-ment* l'emporte sur la suffixation en *-ée*, *-erie*, *-ade*, *-ure*, *-aison*, *-ance*, *-eur*, *-oire*, *-tion*. De même l'inversion en matière de composition de phrase est perçue sur la base de la fréquence et sera classée en tant que fait de style.

4. Au-delà de cette analyse linguistique portant sur la distribution de la construction, se pose dans un premier temps la question de l'existence du mot en termes de confrontation avec un capital lexical (*The mental lexicon*, selon Miller 1991 : 121-143), variable d'un individu à l'autre, représentation (déformée) du capital lexical (variable lui aussi) enregistré par les ouvrages lexicographiques. On peut penser que les mots construits par dérivation sont connus non pas en tant qu'unités mémorisées mais en tant que résultats d'application de règles. La preuve en est que les enquêtes faites sur l'intuition néologique révèlent que des locuteurs de formation équivalente hésitent dans un cas sur deux à reconnaître si le mot dérivé qui leur est présenté est un néologisme (Corbin 1987 : 599-606). *Accrochement* est-il un néologisme ? *L'accrochement, en chimie, désigne une opération qui permet d'associer des molécules en vue de produire des polymères*. Cette phrase — pure invention — suffit pour rendre le mot acceptable. Elle définit le mot par rapport à une langue de spécialité. De même, d'un point de vue pragmatique, la position sociale privilégiée du locuteur suffirait à favoriser l'acceptabilité, s'il ne s'agissait d'un locuteur allophone. La légitimité sociale est contrariée dans notre exemple par la légitimité linguistique.

5. Le jugement d'acceptabilité se trouve, dans un deuxième temps, informé par le recours aux ouvrages lexicographiques. Il n'existe pas de dictionnaires de phrases, mais les dictionnaires de mots foisonnent. On sait qu'ils sont dans l'impossibilité d'enregistrer le lexique dans son extension et dans sa mouvance. Le lexique est infini et se trouve représenté par un choix dans l'ouvrage lexicographique. La datation est souvent approximative, plus particulièrement dans le cas des mots dérivés et on pense aux approximations auxquelles donne lieu l'analyse des dérivations régressives. La non-attestation d'un mot construit par dérivation est souvent accidentelle. Elle fait rarement l'unanimité des dictionnaires. On sait, par exemple depuis l'étude de J. Dubois (1962), que les mouvements de suffixes sont très importants : ils concernent près d'un quart des mots.

Il apparaît donc que le jugement d'acceptabilité repose davantage sur une intuition d'attestation, caractérisée souvent par une sous-évaluation craintive du corpus, que sur une vérification systématique — on n'oserait dire *exhaustive* — de l'attestation lexicographique. *Appointage, applicage, argentage, absoluté, dévirginiser, indevinable, visionneur*, existent-ils ?

Ainsi cette confrontation avec la compétence lexicale du récepteur, le *lexicon filter* de Miller (1991 : 111), relève-t-elle de deux facteurs. Le premier est la connaissance plus ou moins étendue des règles. En situation d'aphasie, des patients qui ont perdu partiellement ou totalement leur capital lexical sont capables de produire des mots par dérivation. Le second est la vérification sur un corpus lexicographique objectif, dont le statut réel, c'est-à-dire lacunaire, est occulté par les instances de légitimation. Alors que toute phrase est toujours nouvelle, l'acceptabilité du néologisme lexical en tant que tel

passera par cette zone floue de l'intuition sous contrôle de la connaissance en tant que représentation du lexique ou en tant que vérification, toujours partielle, de celui-ci.

LA MÉMOIRE DES MOTS

La mémoire des mots, qui s'articule à la grammaire dérivationnelle intériorisée, intervient ici dans sa double signification : les mots ont leur mémoire et les locuteurs ont les mots en mémoire.

La compétence lexicale du locuteur est devenue au cours de ces 25 dernières années de plus en plus synchronique, par suite de la méconnaissance du latin et du grec. Elle passe aujourd'hui par la consultation de plus en plus fréquente de l'attestation et de sa motivation étymologique. La détermination de la base pose un certain nombre de problèmes étant donné les apparences de mots construits (ex. *anthropophage*, *apanage* ou *aréopage*). La construction morphologique ne révèle pas nécessairement le sens du mot : *peigner/peignoir* (différent de *arroser/arrosoir*), *rouerie* ne renvoie pas à *rouer*. Il conviendrait d'étudier systématiquement les faits d'étymologie populaire dans le discours actuel et de dépasser les exemples stéréotypés de *bikini*, *péage*, ou *catamaran*. Les mots ont leur mémoire. Si la relation sémantique entre *herbe* et *herbeux*, *soie* et *soyeux* (N + suffixe → adj, sens de «relatif à N») est transparente, seule la mémoire culturelle du sens étymologique permet d'éclairer *eau*, *aqueux* et *anhydre*. L'absence de prédictibilité se retrouve également compensée par la mémoire dans la construction des termes savants. On peut penser que la connaissance de l'allomorphisme, c'est-à-dire des variations formelles qui interviennent lors de la construction des mots par dérivation (*fleur/floral*, *terreur/terroriste*, *barbe/imberbe*, *traduire/traduction*), repose elle aussi sur la mémoire culturelle, même si de nombreuses régularités ont pu être dégagées (cf. D. Corbin 1987 : 283-340).

L'information historique ne peut donc être négligée, mais il reste à déterminer l'importance qu'il convient de lui donner dans la compétence dérivationnelle. De notre point de vue, elle est seconde.

ACCEPTABILITÉ ET CRÉATIVITÉ

Dans la description qu'il fait du lexique, le linguiste postule, à partir des potentialités morphématiques, des formes que le lexicographe ignore : *ambuler* qui est postulé à partir de *déambulation*, *ambulant*, *ambulatoire*, *funambule* et qui ouvre la voie à un éventuel *ambulation*; *ferrovial* tout aussi «acceptable» que *ferroviaire*, *absolvable* ou autre *buvabilité*. Cette liberté qui donne au linguiste le pas sur le socialement enregistré, le locuteur ne se l'accorde pas, ou on ne la lui accorde que dans des cas précis qui informent le jugement d'acceptabilité. Cette potentialité n'est que trop rarement accordée en didactique lexicale, où le statut du potentiel n'est pas autorisé à limiter le statut du lexicographique.

Nous examinerons trois cas qui doivent permettre de nuancer l'acceptabilité.

1. Le cas de l'enfant. L'âge de la découverte d'un système, l'âge du renforcement et des déments. Le mot d'enfant repose sur une généralisation de règles qui correspond en fait à la recherche bien connue de la sécurité. Comme l'a souligné Miller (1991 : 242), la

capacité morphologique de l'enfant se traduit par la tendance à généraliser les morphèmes producteurs. La même attitude est constatée sur le plan de la sémantique ou plus exactement de la maîtrise du sens par l'enfant. À ce niveau, il apparaît donc bien que la connaissance des règles en tant que telles n'est pas liée à l'étendue du lexique mais que la mémoire des cas vient filtrer cette compétence. La généralisation du suffixe *-age* amène *réparage**; celle du suffixe en *-able* amène *incorrigeable** ou en *-ment* *abandonnement**.

2. Le cas de l'allophone. La même tendance se rencontre chez l'allophone, que représentait au départ notre commissaire européen. Le jugement d'acceptabilité est ici souvent moins souriant. Quelques exemples : *torpide, turpide* (A → N) : *torpitude**, *turpitude*, ou encore : *traduire* (V → A) : *traduisible* ou *traductible*, (A → N) : *traduisibilité* ou *traductibilité*. *Traductible* (cf. *conductible, déductible* ou *réductible*) est absent du *Robert* et est présent dans le *T.L.F.* Or *traductibilité* est présent dans le *Robert*. Seul *traductible* peut expliquer *traductibilité*. Si l'on prolonge l'exemple *traducteur* par rapport à *traduire* (cf. *instructeur* par rapport à *instruire*), on trouve *adducteur/ladduire**. Autant de faits de concurrence ou de lacunes que l'allophone comble aisément, alors que le francophone se paralyse, parce que le prescriptif l'a emporté sur le prédictif au cours de sa scolarité.

3. Le cas du spécialiste. Il s'agit plus ici d'une aire de discours que de catégories de locuteurs. La langue de spécialité permet le déploiement de l'invention parce que le locuteur acquiert à cet endroit une légitimité qui lui est refusée dans les deux premiers cas et pourtant la timidité dérivationnelle y est perceptible si l'on examine la préférence accordée à l'emprunt. À côté des syntagmes complexes et de l'interfixation, qui obéit aux mêmes règles, la construction par dérivation joue un rôle important dans la restriction des propriétés sémantiques et dans son expansion sur la base notamment de sigles lexicalisés. On notera cependant que l'utilisation de mots construits en langue de spécialité tient tout son sens des systèmes construits qu'ils constituent selon les domaines (cf. D. Blampain 1995). À la limite, la mise en forme lexicale contribue à produire «l'illusion de la systématisme et à travers celle-ci la coupure entre le langage spécialisé et le langage ordinaire qu'elle opère, l'illusion de l'autonomie du système.» (P. Bourdieu 1982 : 174). P. Lerat recommande de son côté de ne pas être dupe des «néologismes in vitro» (1995 : 184). Déjà en 1962, J. Dubois montrait comment en chimie, domaine exemplaire de la construction par dérivation depuis Guyton de Morveau et Lavoisier, on avait poussé jusqu'à ses limites le processus de troncature arbitraire. Le traducteur, lui, constamment confronté aux langues de spécialité, n'est pas autorisé à transgresser les règles comme l'auteur, alors que la recherche de l'équivalence stimule sa créativité. En tant que spécialiste de son domaine, il parlera de *traductique* ou de *traductibilité*, mais ira-t-il jusqu'à parler comme M. Serres, dans son ouvrage sur *La traduction*, d'une philosophie de la *duction* ?

De l'observation de ces trois champs de performance et de l'analyse des paramètres du jugement d'acceptabilité, on se sent en droit, pour contribuer à assurer un autre avenir à la langue française, de manifester le souhait que l'on en vienne à dépasser les lois de la reproduction en favorisant le déploiement inconscient de la création morphématique — ou du moins en ne la sanctionnant pas — et en ignorant les frontières des aires légitimes où peut se déployer la création. L'essentiel est de maintenir l'efficacité de la communication et le jugement d'acceptabilité lexicale doit pouvoir relativiser le poids lexicographique. L'image d'un lexique surestimé dans sa légitimité pèse trop sur le jugement

d'acceptabilité. Les schèmes de production sont jugés en français plus que dans une autre langue. La faiblesse de celui-ci à l'échelle internationale passe par l'excès de rigueur. La crainte exacerbée de ne pas trouver les mots légitimés devrait ne pas s'effacer devant la candeur de l'enfant, les trouvailles de l'allophone ou la liberté du spécialiste et peut-être pas uniquement dans les cas de lacunes dérivationnelles ou de concurrence affixale, mais aussi pour répondre à une motivation d'expression, que les *accros* de l'académisme risquent de juger... inacceptable.

RÉFÉRENCES

- BLAMPAIN, D. (1995) : «Langue française et langue de spécialité», Actes du colloque *Langue française et littérature pour francophone : forces centripètes, forces centrifuges*, Liège, Ed. universitaires *Marche romane*.
- BOURDIEU, P. (1982) : *Ce que parler veut dire*, Paris, Fayard.
- CORBIN, D. (1987) : *Morphologie dérivationnelle et structuration du lexique*, Tübingen, Max Niemeyer, 2 vol.
- CORBIN, D. (1997) : «Décrire un affixe dans un dictionnaire», G. Kleiber et M. Riegel (dir), *Les formes du sens*, coll. «Champs linguistiques», Louvain.
- DUBOIS, D. (1962) : *Étude sur la dérivation suffixale en français moderne et contemporain*, Paris, Larousse.
- GUILBERT, L. (1975) : *La créativité lexicale*. Paris, Larousse.
- HANSE, J. et D. BLAMPAIN (1994) : *Nouveau dictionnaire des difficultés du français moderne*, Louvain-la-Neuve, De Boeck-Duculot.
- LERAT, P. (1995) : *Les langues spécialisées*, coll. «Linguistique nouvelle», Paris, Fayard.
- MILLER, G. A. (1991) : *The Science of Words*, New York, Scientific American Library.
- PICONE, Michael (1996) : *Anglicisms, Neologisms and Dynamic French*, Amsterdam/Philadelphia, John Benjamins Publishing Company.
- THIELE, V. et A. CLAS (trad.) (1987) : *La formation des mots en français moderne*, Montréal, Presses de l'Université de Montréal.

UNE MÉMOIRE PLURIELLE OU UNE AMNÉSIE PRODUCTIVE ?

Moufida GHARIANI BACCOUCHE

Université de Tunis I, Tunis, Tunisie

À quelque chose malheur est bon.

Every cloud has a silver lining.

Rubba da:rratin na:fīḥah] «un malheur pourrait être bon».

INTRODUCTION

Si on considère la culture comme la mémoire des peuples, on peut dire que nul autre énoncé linguistique que le proverbe ne serait capable de mieux conserver une telle mémoire.

Ainsi les énoncés proverbiaux seraient-ils les prototypes de séquences dont la fonction mémorielle joue un rôle primordial dans le cadre de la même langue.

En est-il de même des cas où ces séquences ne sont pas rattachables à une seule langue ? Ne s'agit-il pas là d'une mémoire plurielle ou au contraire d'une amnésie productive ?

Dans le premier cas ne serait-il pas question d'énoncés proverbiaux translinguistiques transcendant les différences culturelles pour exprimer des vérités universelles, et dans le deuxième cas d'une absence de mémoire donnant naissance à des séquences qui varient d'une langue à une autre et qui, tout en exprimant une expérience humaine commune, sont bien enracinés dans les spécificités linguistiques de chaque communauté ?

C'est en comparant des énoncés appartenant à l'arabe, au français et à l'anglais que nous allons vérifier de telles oppositions en vue de déterminer l'approche didactique la plus adéquate pour chaque type d'énoncé.

1. UNE MÉMOIRE PLURIELLE

Les énoncés qui nous intéressent dans ce cas se caractérisent par une similarité quasi-parfaite lorsqu'on passe d'une langue à une autre : à l'identité du signifié correspond une ressemblance de la forme.

Sur le plan formel nous pouvons remarquer :

1.1 Une indentité totale

Il y a dans ce cas une ressemblance nette sur les plans à la fois rythmique, syntaxique, morphologique, sémantique et référentiel.

Sur le plan rythmique, on relève les mêmes groupes binaires ou ternaires dans les trois langues. Exemple :

[Al- ʕajnu bil- ʕajni wa-s- sinu bi-s-sinni]
Oeil pour oeil, dent pour dent.
Eye for eye, tooth for tooth.

Le rythme doublement binaire est en effet conservé dans les trois langues.

On peut dire de même des structures syntaxiques de cet exemple ainsi que de tant d'autres comme :

-*Tel père, tel fils.*
- *Like father, like son.*

déterminant + nom/déterminant + nom

Donc : phrase nominale sans verbe, d'où cette concision extrême.

Par ailleurs, il arrive que le signifiant de ce type d'énoncé connaisse la même troncation par apocope (chute d'un segment final), exemple :

- *Chacun pour soi, (et Dieu pour tous).*
- *Every man for himself, (and God for us all).*

À cela, s'ajoutent un signifié unique et une identité référentielle.

Dans le cas du signifié, il arrive souvent qu'on retrouve des définitions lexicographiques rendant compte du même contenu sémantique.

when/while the cat's away, the mice will play.
Quand le chat n'est pas là, les souris dansent. ou /
Le chat parti les souris dansent.

[Y'ib ja: qiṭ ?al ʕab ja: fa:r] (Littéralement «absente-toi, chat; joue souris»)

Ces trois énoncés font tous l'objet de définitions lexicographiques semblables. Comparons la définition française et la définition anglaise :

«Quand les maîtres sont absents les écoliers ou les inférieurs se livrent au désordre»
(*Nouveau Larousse Universel*).

«People will take advantage of the absence of authority, supervision to do as they like, enjoy themselves» (*The Oxford Dictionary of Current Idiomatic English*).

Nous avons ici deux points à préciser :

a) Puisque l'identité lexicale est impossible lorsqu'on passe d'une langue à une autre, ces énoncés proverbiaux représentent des calques parfaits. Exemple :

- *Au royaume des aveugles, les borgnes sont rois.*
- *In the country of the blind, the one-eyed man is king.*

b) Ce dernier point a pour corollaire un fonctionnement symbolique similaire où les mêmes référents symbolisent d'une manière prototypique les situations ou les comportements à désigner. Dans l'exemple :

Quand le chat n'est pas là, les souris dansent,

la relation chat/souris symbolise ou renvoie à toute situation où les attitudes changent en fonction des rapports de force établis en présence ou en l'absence des deux parties, le dominant et le dominé.

Ces ressemblances exceptionnelles ne peuvent être expliquées que par l'adoption des mêmes énoncés, ayant pris naissance dans une langue, par d'autres langues.

Il s'agit donc d'une forme particulière d'emprunt.

Et puisque l'origine anonyme du proverbe est un élément définitoire, l'opposant à la maxime ou à l'apophtegme, il est vraiment difficile de déterminer la langue où ces énoncés proverbiaux étaient initialement conçus. À l'exception de quelques cas très précis dont l'origine est bien délimitée comme la loi du talion (origine religieuse) ou source historique comme dans le cas de l'anglais :

The last straw that breaks the camel's back,

qui vient probablement de l'arabe :

[al qij]ʃaʔallati qaṣamat ṣahra-l-ba ʕiri] (Littéralement «la paille qui a brisé le dos du chameau»).

Le mot «camel» (chameau) étant culturellement marqué, nous pouvons donc dire que la mémoire de ces énoncés ne peut être que plurielle.

1.2 Une identité partielle

Elle ne peut être abordée que sous l'angle d'une différenciation progressive allant d'une simple variation lexicale à des ruptures à la fois syntaxiques, rythmiques et symboliques.

Nous voyons ainsi se dessiner deux pôles, l'un tire du côté des universaux proverbiaux et l'autre du côté des énoncés idiomatiquement bien ancrés dans leurs langues spécifiques.

Au premier pôle figurent des énoncés du type :

- a) - *Quand on parle du loup (on en voit la queue)*
- *Talk of the devil (and he appears)*

- b) - *Tel père, tel fils*
- [Wild -il-fa:r jaʔlaʕ haffa:r]

Dans le premier exemple, il y a substitution d'un item lexical (loup = wolf/devil = demon), avec conservation d'une identité syntaxique (construction corrélatrice), rythmique (groupe binaire) et morphologique (la troncature est possible dans les deux cas).

Dans le deuxième exemple, la différence est plus importante puisque l'équivalent arabe est littéralement :

«Le fils de la souris creuse de naissance des terriers»,

où, tout en conservant la notion de père et fils, on s'en éloigne en ajoutant à la fois la symbolique du rat et celle des terriers.

Aux antipodes de ce groupe figurent les cas où ce sont les dissemblances qui priment.

Pour dire qu'«on néglige souvent les avantages qu'on a immédiatement à sa portée» (*Nouveau Larousse Universel*), on dispose dans les trois langues des exemples suivants :

En arabe, nous avons une série de variantes :

- 1- [Da:ri:l hadda:d bla: sikkina] (Littéralement «La maison du forgeron est sans couteau»)
- 2- [Darin naʕʕa:r bla: ba:b] (Littéralement «La maison du menuisier est sans porte»)
- 3- [Naʕʕa:r wba:b da:ru mʔaf ʔaq] (Littéralement «Menuisier, et la porte de sa maison est délabrée»)

où déjà, dans la même langue, il y a des différences de structures syntaxiques (2° et 3° exemples) et de supports symboliques (forgeron, menuisier, porte, couteau). Si on les compare aux deux énoncés équivalents français et anglais :

Les cordonniers sont les plus mal chaussés.
Who is worse shod than the shoemaker's wife ?

on remarque que le champ symbolique est de plus en plus diversifié (ce qui nous donne les points de dissemblance) même si cette diversité est conçue dans le cadre des professions (d'où les points de ressemblance).

Si on ajoute à cela les différences syntaxiques (phrase nominale en arabe, phrase attributive en français et modalité interrogative en anglais), et par conséquent la structure interne de chaque énoncé, on se rapproche de plus en plus d'énoncés dont le seul point commun est le contenu sémantique.

En effet, chaque langue gère un univers normatif qui lui est propre, d'où «la contrainte du signifiant»; mais nous pouvons également évoquer la volonté des individus

et des groupes de se libérer de ce joug et de créer un espace de liberté et d'identité où le rêve des choses devient une source d'inspiration et d'inventions, où «se ressourcer la langue» par son pouvoir de restituer aux mots de la parole le «pouvoir réminiscent de leurs images» comme l'exprime (Fedida 1985 : 5-20) en ajoutant «cela est possible si toutefois la parole retrouve les métaphores langagières de la langue puisque la langue commune est là toute disponible pour que la parole y trouve langage et ainsi mémoire» et comme le résume si bien A. Rey : «on peut donc poser l'hypothèse d'une prise en charge infiniment variée des structures symboliques d'une culture par les moyens de la langue, moyens lexicaux et syntaxiques» (1977 : 200).

D'où la productivité, la richesse et la diversité des expressions langagières, en l'occurrence, le proverbe qui, selon la formule de Hamm (1989 : 181) devient «pro-verbe, acte de langage par procuration», exprimant des expériences humaines communes, et qui, grâce à son originalité prosodique qui le rapproche de la poésie, est là, tout prêt à être mémorisé pour servir les besoins de la communauté linguistique où il a pris source.

2. UNE AMNÉSIE PRODUCTIVE

Si on reprend la définition que présente Georges Kleiber du proverbe, qui est une dénomination d'un type particulier, si on reprend l'idée d'interdépendance dans les couples proverbe/culture et culture/langue, et si on admet que l'identité culturelle, c'est-à-dire la mémoire profonde des communautés humaines, ne se réalise pleinement que dans l'outil de communication qui est la langue, il est naturel que l'écrasante majorité des énoncés proverbiaux ne retiennent de la mémoire commune que le signifié qui les rattache les uns aux autres d'un côté, et de l'autre côté à cette «sagesse ancienne», ce «prêt à penser», ou cette «moralité en boîte» (Buridant 1984 : x); le signifiant, les items lexicaux, les structures syntaxiques et la charge symbolique demeurent spécifiques aux langues respectives.

N'avons-nous pas là des signifiés qui ont perdu les traces de leur mémoire dans les méandres des spécificités linguistiques et culturelles ?

L'analyse d'échantillons empruntés aux trois langues illustrent en effet la tendance naturelle des langues à ancrer les énoncés proverbiaux dans ce qu'elles ont de plus propre : en fait elles versent des matériaux linguistiques différents dans des moules ou des schèmes sémantico-logiques identiques (cf. Salah Mejri 1997). Si nous prenons les exemples suivants :

- *Tel est pris celui qui pensait prendre.*
- [Man hafara Subban li ?axi:hi waqaSa fi:hi] (Littéralement «Celui qui creuse un puits pour son prochain, tombe dedans»).
- *The biter is sometimes bit,*

nous remarquerons que les trois langues n'optent pas pour le même procédé pour traduire la relation de contrariété établie entre l'actant agissant se transformant, selon le principe énoncé par le proverbe, en actant-patient.

Le français, ici, privilégie la voie hyperonymique, se contentant ainsi d'un verbe relativement générique comme «prendre» et d'une tournure syntaxique énonçant le même procès, soit selon l'angle de l'activité ou de la passivité.

L'arabe et l'anglais optent pour le procédé métaphorique qui passe par le filtre des correspondances symboliques.

En arabe la métaphore est plus compliquée puisqu'elle fait intervenir un procès complexe impliquant l'image du puits, lui-même renvoyant à celle de piège.

Les exemples sont nombreux; on peut en citer d'autres, par exemple : pour dire que personne d'autre que celui qui vit une situation difficile ou pénible ne peut l'apprécier à sa juste valeur, la sagesse populaire française a gardé en mémoire :

Nul ne sait mieux que l'âne où le bât blesse.

Quant à la mémoire anglaise, elle a favorisé le malaise causé par des chaussures étroites en s'exprimant ainsi :

Only the wearer knows where the shoe pinches.

Par contre pour l'arabe, c'est la braise qui fait vivre la plus grande expérience douloureuse :

[Ma:ʒhiss i ʒʒamra ka:n illi jaʕfis ʒliha] (Littéralement «Nul ne connaît mieux la braise que celui qui marche dessus»).

La diversité des choix symboliques, appuyée par les différences lexicales, syntaxiques et rythmiques ouvre le champ grand ouvert devant une productivité infinie, chaque fois différenciée, faisant ainsi de l'amnésie de départ une richesse sur laquelle se bâtissent les spécificités linguistiques.

3. CONCLUSION DIDACTIQUE

Sur le plan purement didactique, les deux groupes déjà dégagés se prêtent à deux approches didactiques différentes :

Le premier type qui rassemble des universaux à la fois linguistiques et idéels pourrait constituer le socle de l'acquisition des énoncés proverbiaux. Il correspond ainsi au premier niveau de l'apprentissage. Le deuxième niveau renfermerait les proverbes partiellement identiques qui obéissent à un classement fondé sur la gradation dans les complications tirées des différences entre les langues.

Par conséquent, plus il y a de différences, plus les formes sont dissemblables, et plus on s'approche du dernier groupe dans lequel ne figurent que les énoncés idiomatiquement marqués.

Une telle approche didactique a le mérite de conjuguer les avantages des deux méthodes, sémasiologique (du signe vers le concept) et onomasiologique (du concept vers le signe) qui ne sont en fait que des manières très proches de celle utilisée dans la description du corpus étudié.

RÉFÉRENCES

Générales

- BURIDANT, Cl. (1984) : *Richesse du proverbe*, Études réunies par Cl. Buridant et F. Suard, Lille, Presses universitaires de Lille III.
- BURIDANT, Cl. (1984) : *Richesse du proverbe*, vol I, «Le proverbe au moyen âge», Lille, Presses universitaires de Lille III.
- CRAM, D. (1983) : «The linguistic status of the proverb», *Cahiers de lexicographie*, XLIII, 2, pp. 53-71.
- FEDIDA, Pierre (1985) : «La résonance atonale. Sur la condition de langage de l'analyste», *Le langage pris dans les mots*, Peuple Méditerranéen, n° 33, pp. 5-20.
- HAMM, A. (1989) : «Remarques sur le fonctionnement de la négation dans les proverbes : l'exemple de l'anglais», *Europhras 88. Phraséologie contrastive*, Ed. Greciano, Faculté des sciences Humaines, Strasbourg, pp. 177-194.
- KLEIBER, G. (1989) : «Sur la définition du proverbe», *Europhras 88. Phraséologie contrastive*, Ed. Greciano, Faculté des Sciences Humaines, Strasbourg, pp. 233-252.
- MAKKAI, Adam (1978) : «Idiomacity as a language Universal», *Universals of Human Language*, Joseph H. Greenberg Ed., Stanford University Press, pp. 401-448.
- MEJRI, S. (1997) : *Le figement lexical*, Publications de la Faculté des Lettres - Manouba, Tunis.
- REY, A. (1977) : *Le lexique : image et modèles. Du dictionnaire à la lexicologie*, Paris, Arnaud Colin.
- RIDOUT, R. and C. WITTING (1969) : *English Proverbs Explained*, Pn Books & Sydney.
- SEITAL, P. (1976) : «Proverb : A social use of Metaphor», *Folklore Genres*, D. Ben Amos Ed., Austin, Université of Texas Press, pp. 125-144.

Dictionnaires

- BAALBAKI, Munir (1982) : *Al Mawrid. A Modern English-Arabic Dictionary*, Dar El Ilm Lil Malayen, Beirut- Lebanon.
- COWIE, Al., MACKIN, R. and Í. MCCAIG (1983) : *Oxford Dictionary of Current Idiomatic English*, VI : Phrasal verbs, V2 : Phrases, clauses and sentence Idioms.
- Harrap's Shorter Dictionnaire Anglais Français / Français Anglais*. (1996) : Chambers Harrap Publishers Ltd.
- Nouveau Larousse Universel* (1949) : 2 tomes.
- SIMPSON, John (1982-1988) : *The Concise Oxford Dictionary of Proverbs*, Oxford University Press.

LES MOTS SUR LE DIVAN

Alexis NOUSS

Université de Montréal, Montréal, Canada

mémoire — mes moires...

mots — à modeler, mouler, meuler.

Michel Leiris, *Langage tangage* ou *Ce que les mots me disent*.

Michel Leiris regroupe en 1969 un ensemble de recueils de textes dans un volume qu'il intitule *Mots sans mémoire*. Facétie traduisant une revendication du poète à réinventer pour chaque énoncé les mots du langage ? Sans doute. Leiris est trop familier des territoires de l'imaginaire pour ignorer la part d'ombre qui non seulement accompagne toute expression langagière mais encore qui la fonde. En outre, le procédé surréaliste qui est le sien dans «Glossaire : j'y serre mes gloses» (1982 : 71-116)¹ consiste à donner des définitions où les mots gagnent leur sens de ce qu'ils peuvent précisément rappeler d'autres mots. Pas de meilleur exemple pour saisir le jeu et ce qu'il indique de la pensée du langage opérante que la définition du terme : «LANGAGE — bagage lent, langes de l'esprit.» (1969 : 96)

La mémoire des mots, dans le contexte freudien, est une notion qui apparaît évidente, principielle et fondatrice, à deux niveaux. La psychanalyse, d'une part, se fonde sur le fait que les mots ont une mémoire, et que cette nature détermine sa fonction essentielle quant aux processus psychiques. Le langage se retrouve en défaut de dénotation mais gagne dans cette déféctuosité les ressources de sa signifiante. D'autre part, dans l'appareil lexical que Freud forge pour rendre compte des phénomènes psychiques, les mots acquièrent la technicité requise sans abandonner la chair mémorielle de leur étymologie ou de leur usage populaire.

Langage et métalangage se rejoignent ainsi dans l'œuvre et la pensée freudiennes, scellant notamment leur adéquation sur la problématique de la mémoire des mots. Les traductions de Freud sont éminemment révélatrices à cet égard, le plus souvent par leurs faiblesses, en ce qu'elles dissocient le contenu de la forme, bafouant le message même du freudisme². Au demeurant, la mémoire est centrale pour le projet psychanalytique qui peut être défini comme une tentative de retrouver les fondements psychiques de l'individu, de

¹ Le recueil date de 1939. Leiris lui donne une suite avec *Langage tangage* ou *Ce que les mots me disent*, 1985, Paris, Gallimard.

² Sur cette question, voir *Meta*, vol. 27, n° 1, 1982, *L'écrit du temps*, n° 7, été 1984, et Ornston, 1992.

l'humanité, de la culture. On a pu à juste titre avancer que pour Freud l'inconscient est avant tout une mémoire et que les deux instances fonctionnent selon les mêmes règles et articulations³. La mémoire des mots est déjà là : en allemand penser se dit *denken* et mémoire *Gedenken* (ou *Andenken*) et *Gedächtnis* (du participe passé, *gedacht*, de *denken*).

1. LE LANGAGE SELON FREUD

Le présent exposé s'inscrit aisément dans le cadre francophone de ces rencontres. C'est en effet la psychanalyse française, notamment dans le sillage lacanien, qui a donné une lecture spécifiquement linguistique à l'œuvre freudienne⁴. Traiter cependant de la question générale du langage en psychanalyse déborderait notre propos et je m'appliquerai à la seule considération des mots dans leur charge mémorielle à partir de trois ouvrages de Freud, en succession chronologique, portant respectivement sur les rêves, sur les oublis de mots et lapsus, enfin sur les mots d'esprits. Il est d'emblée important de lever une objection possible, à savoir que les mots seraient toujours soumis à une logique mémorielle de par leur dimension référentielle : ils rappelleraient le réel au sens qu'ils le désigneraient. Mais la désignation n'est qu'un rappel dans une acception relâchée du terme⁵. Quand je parle de mémoire ici, c'est dans ce que la notion implique de travail au sens freudien, c'est-à-dire de transformation, d'interprétation, de reconstruction, l'opération proprement traductrice que l'on constate dans les élaborations produites par l'activité onirique, premier matériau étudié par Freud et qui lui sert de terrain d'énonciation de la théorie analytique⁶.

L'interprétation des rêves introduit directement le thème qui nous occupe. Le rôle de la mémoire est souligné dès le premier chapitre : «Tout le matériel qui forme le contenu du rêve provient d'une manière quelconque de notre expérience vécue : il est donc reproduit ou remémoré dans le rêve.» (1993 : 18). Le titre⁷ même relève d'un effet de mémoire

³ C'est aussi par sa méthodologie que la psychanalyse est une entreprise de mémoire : dans ses ouvrages, Freud cite systématiquement les travaux antérieurs à ses recherches dans le domaine abordé. Par ailleurs, il fait constamment appel aux monuments livresques de la culture occidentale (la Bible, Sophocle, Shakespeare, Goethe) pour étayer ou illustrer ses thèses. Dans une Vienne et une Europe qui allait succomber à la barbarie, on pourrait considérer son œuvre comme la tentative désespérée de sauver et rappeler cet héritage. La naissance et le développement de la psychanalyse s'inscrivent dans un cadre historique et les totalitarismes qui se mettaient en place s'annonçaient comme des machines d'amnésie. Sur ce point voir Bettelheim, 1993 et Goldschmidt, 1988.

⁴ Il faut également mentionner, quant à cette orientation, l'apport de P. Ricoeur, présentant le freudisme comme une herméneutique : «Ce n'est point le désir comme tel qui se trouve placé au centre de l'analyse, mais bien son langage.» (1965 : 15).

⁵ Quoique d'un point de vue philosophique la référentialité puisse être considérée comme une mémoire. De même que l'exercice du langage est un pari sur la possible disparition du réel (user d'un signifiant, c'est faire le deuil du signifié, affirmer son absence) — dire «table», c'est pouvoir le dire lorsqu'aucune table n'est à portée —, il serait la constante mémoire de ce réel : dire le mot, c'est (se) rappeler que l'objet existe.

⁶ La traduction peut être considérée comme le processus paradigmatique de l'ensemble des opérations propres à la fois à l'économie psychique et à l'activité analytique. Voir Mahony, 1982 et 1992 et Benjamin, 1989.

⁷ J'en conserve la version canonique alors qu'elle révèle un écart traductif : *Die Traumdeutung* devrait se traduire *L'interprétation du rêve*. Le passage du singulier au pluriel n'est pas sans conséquences théoriques et révèle une certaine manipulation.

puisqu'il renvoie aux manuels populaires d'interprétation des rêves, du type *La clé des songes*. Ce choix marque à la fois la reconnaissance d'une vertu heuristique accordée à un tel matériau privilégié par la culture populaire et, par contraste, la radicale différence épistémologique dans son traitement.

Parallèlement, la nature langagière des productions oniriques — sa prise en considération dans l'analyse du rêve constitue l'originalité de Freud — est soulignée. «Il ne faut pas s'étonner du rôle que joue le mot dans la formation du rêve. Le mot, en tant que point nodal de représentations nombreuses, est en quelque sorte prédestiné aux sens multiples, et les névroses (les obsessions, les phobies) utilisent aussi hardiment les possibilités de condensation et de déguisement que le mot présente.» (1993 : 293) Les nombreux récits de rêves et leurs analyses qui parsèment l'ouvrage offrent de multiples exemples de la place centrale du mot dans le travail onirique, indiquée par une typographie en italiques. Or ce rôle n'est pas de nature expressive dans le contexte de la situation rêvée mais il sert de relais mnésique à une situation antérieure⁸. Sa fonction est essentiellement et strictement de mémorialisation : il fixe un souvenir pour l'introduire dans le scénario onirique. Décalage énonciatif structurel qui se joue dans la dynamique de *différance* théorisée par Derrida. Le mot fait passer en contrebande un sens volontairement oublié : un os *brisé* rappellera un adultère (*Ehebruch*, rupture de mariage, en allemand) (1993 : 350); le mois de *juillet* évoquera la pièce *Jules César* et le thème de l'ambivalence affective (1993 : 362).

Dans l'élaboration du contenu onirique et fantasmatique, les mots véhiculent la mémoire d'affects préexistants à la scène rêvée et supportant sa signification profonde⁹. Le rêve dit «de la monographie botanique» tourne autour d'un opuscule consacré à une certaine plante. Dans le récit qui en est fait, entre autres occurrences botaniques, le rêveur croise le Pr. Gärtner (jardinier en allemand) accompagné de sa femme à laquelle il trouve une mine «florissante»; il s'entretient avec lui d'une malade répondant au nom de Flora. La cohérence du réseau lexical floral ainsi mis en place et en scène dépasse visiblement la seule contingence ou l'expression d'une tendance bucolique chez le rêveur mais sert à rappeler des expériences liées à une figuration florale qui, à leur tour, se rapportent aux pensées constituant le contenu latent du rêve¹⁰.

Dans le célèbre «rêve de l'injection faite à Irma», le signifiant «triméthylamine» (d'autant plus important qu'il sert à Freud pour désigner ce rêve) affiche explicitement sa valeur de mot puisque Freud en voit la formule imprimée en caractères gras (le mot l'est également dans le récit) «comme si on avait voulu la faire ressortir tout particulièrement» (1993 : 108) précise-t-il. Or la substance chimique injectée ne figure pas ici pour sa vertu thérapeutique mais parce qu'elle rappelle ce que Fliess, un des interlocuteurs privilégiés de Freud, lui avait dit de la présence du produit dans le métabolisme sexuel. Il analyse :

⁸ Les mots servent aussi de «chaînon intermédiaires», de relais internes, entre les différentes composantes mnésiques du rêve; voir *infra* l'exemple du rêve de la monographie.

⁹ Un effet de cette surdétermination mémorielle du mot est que l'énoncé onirique en devient, d'une certaine manière, idiolectal. Non certes au niveau lexical — il est façonné de la mémoire des autres — mais syntaxique : «Le rêveur invente sa propre grammaire.» (Derrida, 1979 : 310). Le mot ne prend sens que dans le contexte particulier du rêve, dessiné par les réminiscences du rêveur.

¹⁰ Au demeurant, lorsque Freud reprend pour la seconde fois l'analyse de ce même rêve plus loin dans l'ouvrage, au chapitre «Le travail du rêve», son vocabulaire se fait très précis : «le mot *botanique*», «le mot *artichaut*» (pp. 245-246).

«Je devine pourquoi la formule de la triméthylamine a pris tant d'importance. Elle ne rappelle pas seulement le rôle dominant de la sexualité, mais aussi quelqu'un à qui je songe avec bonheur quand je me sens seul de mon avis.» (*id.*)

La nature mnésique du mot dans l'énoncé onirique ressort particulièrement des cas où il joue sur (et se joue des) frontières linguistiques¹¹. Une patiente rêve de sa fille étendue morte dans une boîte, ce qui lui rappelle une discussion, la veille, autour du mot *box*, qu'elle relie à l'allemand *Büchse*, qui sert dans la langue vulgaire à désigner les organes sexuels féminins (1993 : 140).

Un dernier exemple démontrera à l'extrême cette dimension mémorielle puisque non seulement le mot n'y exerce qu'une fonction de rappel mais de surcroît n'a été formé qu'à cette fin. Dans un décor maritime et militaire, en conclusion du récit, le rêveur (Freud) et son frère s'exclament à la vue d'un navire coupé à sa moitié et chargé d'objets évoquant des timbales ou des boîtes : «Voici le navire du petit-déjeuner ! (*Frühstücksschiff*)» (1993 : 396). Freud analyse l'image de ce navire, à l'aide du souvenir d'une antiquité étrusque et d'un vers de Schiller, comme se rapportant aux idées de mort, de naufrage, de deuil et de toilette funéraire. Il précise : «*Frühstück* = petit déjeuner = *breakfast*, *Fasten-brecher* (brise-jeûne). L'idée de briser nous ramène au *naufrage*, l'idée de *jeûne* à la toilette noire.» (1993 : 397)¹².

Parmi les catégories de phénomènes étudiées dans la *Psychopathologie de la vie quotidienne*, celle des oublis de mots (noms propres et communs) nous retiendra d'emblée car l'acte psychique en question révèle un trop-plein de mémoire davantage qu'une lacune ou une faiblesse de la faculté mnémonique. Le mot n'a pas disparu du système psychique, comme effacé par une gomme magique; il est au contraire, présent, trop présent, mais la situation, le sentiment ou la pensée qu'il rappelle ont été refoulés — donc inscrits encore plus profondément —, et l'ont entraîné dans leur opération de dissimulation. «L'action visible» de ces éléments «se manifeste précisément par l'oubli.» (1996 : 29) En compagnie masculine, une jeune fille ne retrouve pas le titre d'un roman récemment lu : il s'agit de *Ben-Hur* dont le second signifiant est proche de *Hure*, prostituée en allemand. «Sans s'en rendre compte, elle considère l'énoncé du titre *Ben-Hur* comme équivalent à une invitation sexuelle et son oubli correspond à une défense contre une tentation inconsciente de ce genre.» (1996 : 52). Autre exemple où l'élément oublié est pourtant rien de moins qu'évident : Freud ne parvient à nommer une station thermale qu'il connaît bien. Le nom retrouvé : *Nervi*. Et Freud de commenter : «C'est que *Nervi* sonne comme *Nerven* (nerfs), et les nerfs constituent l'objet de mes occupations et préoccupations constantes.» (1996 : 31).

¹¹ Dans le premier chapitre, déjà mentionné, où Freud indique la dimension mémorielle du rêve, il rapporte l'observation selon laquelle «les langues étrangères sont souvent parlées d'une manière beaucoup plus pure et beaucoup plus courante en rêve que pendant la veille» (1993 : 19, note 1).

¹² Il y a lieu d'indiquer ici l'importance du facteur de «figurabilité» (et donc de concrétude; voir *infra*) dans les mots des rêves, à savoir la facilité de représentation par images (1993 : 292). Mais cela ne saurait déranger le linguiste saussurien pour qui le signe allie un concept et une *image* acoustique. La dimension mémorielle du mot pourrait être rattachée à sa dimension figurative.

Le mot en tant que trace mnésique apparaît également dans d'autres catégories de constructions symptomatiques¹³. Le «souvenir-écran» participe à l'évidence de cette dynamique puisque sa fonction est de produire un souvenir destiné à en rappeler, par occultation, un autre. Freud rapporte le cas d'un interlocuteur prétendant ne pas se souvenir des grossesses nombreuses de sa mère mais évoquant son image en train de défaire sa jupe. «Or le délaçage (*Aufbinden*) de la jupe n'apparaît dans ce cas que comme un "souvenir-écran" pour accouchement (*Entbindung*). Il s'agit là d'une sorte de "pont verbal" dont nous retrouverons l'usage dans d'autres cas.» (1996 : 62). Le lapsus où un énoncé verbal prend la place d'un autre est de même traité par Freud comme la manifestation d'un élément refoulé ou interdit mais rappelé précisément par la perturbation énonciative. Une patiente remarque à propos de sa famille qu'«ils sont tous avares (sie haben alle *Geiz*; littéralement : ils ont tous de l'*avarice*)... je veux dire : ils ont tous de l'*esprit* (*Geist*).» Tel était en effet le reproche qu'elle avait refoulé de sa mémoire. Or, il arrive souvent que l'idée qui s'exprime dans le lapsus est précisément celle qu'on veut refouler.» (pp. 77-78).

Le refoulement, notion cardinale dans le système freudien, tient, contrairement aux apparences, de la mémorisation et non de l'amnésie. Soumis à son action et y trouvant sa cause, le lapsus n'est pas qu'un phénomène diachronique où un élément prendrait la place d'un autre par contiguïté mais un processus à comprendre synchroniquement qui remplace un signifiant par un autre qui lui étant postérieur pour mieux conserver le premier. Erreurs de lecture et d'écriture¹⁴ participent du même mécanisme : «*Sprachstrategie*» (stratégie linguistique) au lieu de «*Schachstrategie*» (stratégie d'échiquier) lit un philologue à la suite d'une polémique avec des collègues (1996 : 131); le psychanalyste Brill, au lendemain d'une soirée arrosée, remplace le nom d'une malade «*Ethel*» par «*Ethyl*» (1996 : 141).

On ne s'étonnera pas enfin de voir évoqué *Le mot d'esprit et sa relation avec l'inconscient*¹⁵ puisqu'il s'agit de l'ouvrage majeur consacré directement à la question du langage dans les processus psychiques, où Freud montre que ceux-ci sont repérables dans le mot d'esprit au même titre (et pour les mêmes raisons stratégiques) que dans les phénomènes précédents.

Une première grande typologie distingue les mots d'esprit fondés sur des pensées de ceux produits par des manipulations linguistiques. «Les mots sont un matériau plastique avec lequel on peut faire toutes sortes de choses.» (1992 : 87). Souplesse morphologique et sémantique qui revêt divers aspects dont, pour ce qui nous concerne, redonner à un mot l'empreinte d'une signification pleine désormais affaiblie. Tels les exemples suivants :

¹³ La moitié du volume est consacrée à des actes symptomatiques dont la manifestation est verbale, ce qui situe à bon droit la *Psychopathologie de la vie quotidienne* parmi les ouvrages de Freud traitant du langage. Contestant l'explication réduisant la cause des lapsus à des phénomènes strictement linguistiques (contacts des phonèmes), Freud remarque d'ailleurs : «Mais dans les cas où le trouble est occasionné par un élément extérieur à la phrase ou au discours qu'on est en train de prononcer, il s'agit avant tout de rechercher cet élément, et la question qui se pose alors est de savoir si le mécanisme d'un tel trouble peut nous révéler, lui aussi, les lois présumées de la formation du langage.» (1996 : 68). Voir aussi p. 96.

¹⁴ Cette dernière catégorie ne saurait nous étonner, compte tenu de la visée thérapeutique de la psychanalyse : le mot qui désigne un lapsus *calami* dans le lexique freudien, *verschreiben*, signifie aussi prescrire, bel exemple de mémoire du mot.

¹⁵ L'homonymie ne doit pas prêter à confusion : «mot d'esprit» traduit *Witz* mais n'est qu'une solution insatisfaisante et réductrice; voir la «Note liminaire» (1992 : 33).

«"Comment ça marche ?", demande un jour l'aveugle au paralytique. "Comme vous le voyez", répond le paralytique à l'aveugle.» (1992 : 87). À Napoléon I^{er} dénigrant ses compatriotes lors d'un bal de cour en affirmant : «*Tutti gli Italiani danzano si male*» (Les Italiens dansent tous vraiment mal), une dame italienne répliqua : «*Non tutti, ma buona parte*» (Pas tous, mais une bonne partie) (1992 : 82). L'investissement mnésique ne relève cependant pas seulement de la mémoire de la langue. L'économie psychique individuelle intervient là aussi, comme dans les rêves et autres formations symptomatiques, à la différence que si les mécanismes sont similaires, les motivations se situent «en des lieux du système psychologique» (p. 321) distincts. Le mot d'esprit a pour fonction de lever ou détourner une inhibition mais il le fait dans un cadre social, devant un nécessaire auditoire, développant la dimension relationnelle de l'exercice du langage. Les censures ainsi trompées concernent les refoulements et résistances liés à la sexualité ou aux formes du pouvoir (pulsions de lubricité ou d'hostilité, selon les termes de Freud). Le mot, tout en le transgressant, fonctionne comme un rappel de l'interdit. En en restant aux mots d'esprit fondés sur des jeux verbaux, citons pour la première catégorie : «Cette jeune fille me fait penser à Dreyfus. Les militaires ne croient pas à son *innocence*.» (1992 : 96). Pour la seconde, deux exemples rapportés en français : à propos de Napoléon III commençant son règne en confisquant les biens de la Maison d'Orléans, «C'est le premier vol de l'aigle» (1992 : 91); à Louis XV demandant à un courtisan de faire un mot d'esprit dont lui-même, le souverain, serait le sujet : «Le Roi n'est pas un sujet.» (1992 : 92).

Pour Freud (1992 : 235 *sq.*), le plaisir — à entendre comme gain dans l'économie psychique — attaché à l'exercice du mot d'esprit retrouve le plaisir de l'enfant à jouer avec les mots comme alliages de sons et de rythmes sans se soucier du sens et de «la pression de la raison critique». À son tour, ce plaisir est lié à celui des premiers jeux érotiques du tout-petit, avec son corps et celui de la mère. Si bien que le plaisir du mot d'esprit rappelle ultimement le plaisir de l'*infans*, celui qui ne sait pas parler. Dans ces jeux de mots-là, le langage garde la mémoire du non-langage.

2. LE LEXIQUE DE FREUD

Après la langue *pour* Freud, la langue *par* Freud. On l'a désormais suffisamment souligné, la langue des philosophes n'est pas plus transparente ou innocente que celle des poètes. Le constat est évident pour Nietzsche, Marx ou Heidegger mais pas moins pour Descartes ou Spinoza. En outre, Foucault, s'interrogeant sur la notion d'auteur, introduit à propos de Marx ou de Freud la notion de «fondateurs de discursivité» qui, à la différence du simple écrivain, produisent, au-delà de leurs œuvres, «la possibilité et la règle de formation d'autres textes [...], une possibilité indéfinie de discours» (1994 : 804-805). Une telle ouverture est fournie par l'établissement d'un champ épistémologique balisé de concepts originaux et de procédures réflexives particulières. Mais il ne saurait être efficacement instauré s'il n'était pourvu d'un appareil linguistique qui, plus qu'une terminologie, s'apparente à une langue en ce que la structuration d'une langue — morphologie, vocabulaire et syntaxe — permet une reprise et une créativité sans limites. Le constat est d'autant plus pertinent dans le cas de Freud que la démarche psychanalytique s'appuie précisément sur les fonctions du langage, pour l'élaboration du matériau à traiter aussi bien que pour son interprétation¹⁶.

¹⁶ Sur Freud et sa langue, voir les travaux de P. Mahony et J.-M. Rey (voir références).

La langue forgée par Freud pour établir le champ et le discours psychanalytique fait appel à la dimension mnémonique de l'exercice langagier sous divers aspects. C'est l'allemand quotidien, utilisé au maximum de ses possibilités de multisémantisme et de dépendances contextuelles, une langue qui garde les traces de ses emplois et significations, une langue qui garde mémoire de la rue même lorsqu'elle prend demeure dans le cabinet de l'analyste ou la bibliothèque de l'écrivain : *der Trieb*, la pulsion, ne doit faire oublier que le mot dit aussi le penchant, la poussée, la force motrice; *die Verdrängung*, le refoulement, presse et pousse; *das Unbewusste*, l'inconscient, affirme tout simplement : ce qui n'est pas su et le souvenir, *die Erinnerung*, ce qu'on a mis à l'intérieur de soi; *der Zwang*, la compulsion ou contrainte, garde la marque de la force physique (au sens de : être forcé à); *der Einfall*, la libre association, plus qu'elle ne compose ou ne relie, témoigne d'une irruption; *die Hemmung*, l'inhibition, c'est aussi l'arrêt, le ralentissement ou le freinage; *die Versagung*, le déni, ne refuse pas le dire, elle l'affronte. Quant aux mots que la traduction anglaise déguise en latin, *ego* et *id*, ils conservent en allemand, *das Ich* et *das Es*, l'affectivité de pronoms familiers.

La néologisation prend également chez Freud des voies directes, sans passer par les langues mortes, effet de mémoire propre à la langue dans laquelle il pensait et s'exprimait. La néologisation en allemand ne procède pas par innovation et découverte mais par rappel et transparence. Grâce à ses possibilités d'autocomposition et d'expansion par juxtaposition d'éléments de la langue usuelle¹⁷, l'allemand retient le dynamisme d'une parole vivante sans devoir recourir à la médiation d'une étymologie savante et aliénante. *Die Fehlleistung*, l'acte manqué, se compose de *Leistung*, l'accomplissement, la réalisation, et *fehl* qui indique le ratage, l'erreur; *das Mischwort*, le mot mixte ou mot-valise, adjoint au mot «mot» le radical *misch-*, au sens de mélange; *die Deckerinnerung* : souvenir-écran ou souvenir-couverture; *der Seelenapparat*, l'appareil de l'âme ou appareil psychique. Tout locuteur allemand se sent en territoire connu même si encore inabordable.

Diverses motivations à ce recours au parler commun : «Les mots de notre langue quotidienne ne relèvent de rien d'autre que d'une magie affaiblie¹⁸». Tout en adoptant une vigilance critique jamais en défaut quant à ce que les mots veulent dire au-delà de leur prétention sémantique¹⁹, Freud ne se départit pas d'une confiance dans l'authenticité des matériaux charriés par les traditions populaires (la langue, les légendes, les croyances)²⁰. Par ailleurs, Mahony, pour expliquer l'utilisation de l'allemand quotidien par Freud, s'appuie sur une remarque avancée quant au matériau servant au travail du rêve : «[...] en toute langue, les termes concrets, par suite de leur évolution, présentent plus de points de contact [pour les associations possibles] que les concepts.» (1993 : 292). Langue concrète qu'est au demeurant l'allemand, «presqu'incapable d'abstraction» selon Goldschmidt (1988 : 31) qui souligne la spatialité, la dynamisation et la corporalité qui structurent son

¹⁷ Il nous faut aussi mentionner la richesse créatrice des préfixes : *ver-*, *un-*, *über-*, *auf-*, *ab-*... ainsi que les innombrables variations à partir de *-setzen* ou *-stellung*.

¹⁸ «*Die Worte unserer täglichen Reden sind nichts anderes als abgeblasster Zauber.*» cité par D. G. Ornston Jr. (1992 : 64).

¹⁹ Sur ce point, Freud appartient bien à la Vienne du tournant du siècle, rejoignant dans cette inquiétude les grands agitateurs de la crise du langage propre à la modernité, Hofmannstahl, Kraus ou encore Wittgenstein.

²⁰ Dans *L'invention du quotidien*, essai «dédié à l'homme ordinaire» (1990 : 11), M. de Certeau s'appuie d'emblée sur Freud, certes dans ses derniers ouvrages mais cela ne saurait effacer le souci affirmé d'une portée universelle à ses thèses dès le début de l'œuvre.

lexique, concrétude qui se poursuit dans la langue des philosophes²¹, contrairement à sa réception dans le domaine français. Dans «concept» en allemand, *der Begriff*, il y a la prise, la poignée ou encore... la griffe. L'«essai» des *Trois essais sur la théorie sexuelle*²² — que Freud rédigea en même temps que *Le mot d'esprit et sa relation à l'inconscient* — se dit en allemand *Abhandlung* où se reconnaissent la main, l'action et même le commerce.

Le souci d'intelligibilité dans la langue de Freud part pour Bettelheim de la visée essentiellement humaniste de l'entreprise psychanalytique. Ce qui est énoncé doit renvoyer le lecteur à son propre destin, le rappeler à sa propre identité : «Freud, dans le choix de ses mots, et par le biais de son style direct, sert bien son dessein de faire en sorte que le lecteur applique à lui-même ses connaissances psychanalytiques, parce que c'est à partir de son expérience intérieure qu'il peut vraiment comprendre ce dont parlent les textes de Freud.» (1993 : 65) «Freud évitait les termes techniques ésotériques [...] parce que la psychanalyse veut, par essence, faire connaître ce qui est inconnu, rendre accessibles et compréhensibles des idées communément ignorées.» (1993 : 176). La langue de Freud, gardant la mémoire de la langue de ses lecteurs, les renvoie à leur propre mémoire.

Rien ne le montrerait davantage selon Bettelheim que l'emploi persistant du mot âme, *die Seele*²³, pièce de résistance de son livre, comme l'indique le titre *Freud and Man's Soul*. Il montre par de nombreux exemples que ce que la *Standard Edition* traduit en anglais par le réseau lexical esprit-intellect-mental est exprimé chez Freud par ce terme qui, par métaphore — on ne saurait défaire l'athéisme radical de l'auteur de *L'avenir d'une illusion* —, illustre le siège à la fois de l'intelligence et des passions, englobe les niveaux conscient et inconscient et leurs relations, terme volontairement ambigu qui reflète l'ambiguïté du psychisme, «l'essence même de notre humanité» (1993 : 159). La centralité du signifiant affirme une continuité culturelle et un effet de mémoire : «En évoquant l'image de l'âme, Freud met l'accent sur le fonds d'humanité que nous avons en commun.» (1993 : 151) et il cite à l'appui la remarque lexicologique suivante de Freud : «"Psyché" est un mot grec et sa traduction allemande est "âme". Donc traitement psychique veut dire "traitement de l'âme".²⁴» (1993 : 154). Bettelheim, fort de sa proximité linguistique et culturelle avec l'univers au sein duquel naquit l'œuvre freudienne, fut l'un des premiers auteurs à insister sur la nature de la langue de Freud et à relever les

²¹ «Le chapitre I de la *Phénoménologie de l'esprit*, "La certitude sensible" [...], est du premier au dernier mot fait avec le vocabulaire qu'un enfant de cinq ans a à sa disposition (à l'exception peut-être des mots *Vermittlung*, médiation, et *Unmittelbarkeit*, immédiateté).» (Goldschmidt, 1988 : 17).

²² À la première page, lorsque Freud introduit la notion de pulsion sexuelle, il le fait en déplorant une lacune : «Il manque au langage populaire une désignation équivalente au mot "faim"; la science emploie à cet effet le terme de "*libido*".» (1995 : 37).

²³ Que pour sa part Goldschmidt (1988 : 149) rapproche de *die See*, la mer, thème central de son livre et dont la métaphorisation éclaire sa réflexion sur la langue de Freud : «[...] c'est que l'inconscient risque bien, en effet, de n'être rien d'autre que cette "voix" qui parle dans la langue et que, justement, les paroles "couvrent", de même que la surface empêche de voir le fond mais en est la remontée, l'émanation. Seule la surface me montre qu'il y a le fond. Tout ce qu'a écrit Freud était déjà comme un possible de la langue.» (1988 : 150).

²⁴ «"Psyche" ist ein griechisches Wort und lautet in deutscher Übersetzung Seele. Psychische Behandlung heisst demnach Seelenbehandlung.»

inconsistances des traductions²⁵. Conserver les strates signifiantes du texte freudien, y compris les imprécisions et les flottements sémantiques, ne pas les neutraliser, les dénaturer ou les nationaliser revient à conserver la mémoire des mots, impératif traductif s'il en est.

D'autres aspects de la langue freudienne sont à retenir dans notre perspective. C'est aussi dans sa spécificité une «langue de spécialité» constituée à partir d'autres spécialités, d'autres lexiques disciplinaires : médecine, physique, mécanique, philosophie, et, qui, ne reniant pas ses sources, en fonde sa légitimité puisque la psychanalyse entend jeter une lumière sur le fonctionnement global de l'esprit humain²⁶. C'est encore une langue qui sans être érudite véhicule une connaissance de la culture classique (la mythologie grecque par exemple) que Freud suppose partagée par des lecteurs ayant été formés aux mêmes sources. C'est enfin une langue affichant une grande maîtrise stylistique, une langue d'écrivain dont les qualités littéraires ont été saluées aussi bien par Thomas Mann que Hermann Hesse ou Albert Einstein. En ce sens, elle n'opère pas de rupture dans la tradition littéraire allemande mais s'inscrit dans sa continuité, comme en témoigne le constant attachement de Freud à la figure et à l'œuvre de Goethe²⁷.

Littéraire, la langue freudienne l'est par son style pour Mahony qui en a étudié les dimensions créatrices et exploratoires : spontanéité, ambiguïté, expressivité, fluidité, dialogisme, figuralité, métaphorisation, un «langage suggestif et associatif» (1990 : 250) qui, mêlant la rationalité et l'affectivité, la démonstration et la description, devient «une écriture-frontière qui oscille entre le conscient et l'inconscient» (1990 : 60)²⁸. Le lecteur est ainsi invité à expérimenter dans sa lecture la dynamique d'association libre propre à l'analyse et le texte freudien suscite une «lecture psychanalytique» (1990 : 231 sq.)²⁹. Langue des résonances et des réminiscences dont Freud infuse son texte et auxquelles le récepteur est appelé à répondre par les siennes.

Loin des lourdeurs à la fois scientificisantes et sacralisantes souvent adoptées par ses traducteurs³⁰, l'allemand de Freud se donne à une compréhension vivante et directe, revendiquant une fidélité pragmatique aussi bien que culturelle : les mots y gardent leur mémoire, ce qui sied à une méthode dont la visée est de remonter à la vérité d'une subjectivité en en suivant les traces mnésiques. La langue de Freud ainsi considérée

²⁵ Sa critique ne manqua pas d'être critiquée. Les griefs majeurs retenus sont en général son dogmatisme, sa méconnaissance des usages diachroniques et multiples de l'anglais et son passéisme idéologique et culturel. Voir Ornston (1992) et F. Kermodé. «Freud is Better in German» in *The Uses of Error*, London, Fontana Press, 1991.

²⁶ «Emprunt, empreinte, emprise» seraient selon M. Schneider les modalités de ce transfert qui s'attache aux mots comme aux idées (1985 : 137).

²⁷ Entre autres références majeures, dont Schiller et Lessing.

²⁸ Ornston (1990 : 68) rappelle que Freud se refusait aux définitions claires et définitives afin de ne pas trahir le matériau inconscient.

²⁹ Voir aussi Mahony (1992) et *On Defining Freud's Discourse*, New Haven and London, Yale University Press, 1989.

³⁰ L'entreprise de traduction des *Œuvres complètes* de Freud en français par l'équipe de Jean Laplanche a suscité discussions et controverses. Faute de pouvoir ici l'examiner, indiquons pour notre propos qu'elle mêle une néologisation parfois précieuse à l'utilisation parfois archaïsante des ressources du français.

pourrait servir à alimenter deux débats de grande importance dans l'épistémè contemporaine.

Le premier touche à la question de l'authenticité. J'emprunte le terme à Adorno qui dans son essai sur Heidegger (1989)³¹ dénonce l'utilisation d'une langue qui, dans sa revendication d'une pureté ou d'une purification ontologique, en vient à créer une mythologie conceptuelle et un fétichisme lexical aliénant l'humain du seul cadre où se déploie et se joue son humanité, l'histoire. Les termes du débat sont clairs : peut-on parler de l'humain en se coupant de son langage ? À quoi vaut de viser l'Être si le sujet s'y perd³² ? Freud répondait : «Vous allez sans doute trouver mauvais que nous ayons choisi, pour désigner deux instances ou provinces psychiques [le moi et le ça], des mots courants au lieu de vocables grecs sonores. Mais nous aimons, nous autres psychanalystes, rester en contact avec la façon de penser populaire et préférons rendre utilisables pour la science des notions populaires que de les rejeter.» (1970 : 110).

Le second débat concerne la légitimation scientifique et a souvent été abordé en regard du développement des sciences humaines. La scientificité cherche à être garantie par son expression : une langue scientifique se veut sans mémoire car elle fonde sa valeur sur la nouveauté de son apport épistémologique et sur l'adéquation à son objet. Quoi de mieux pour le prouver que la nouveauté de son vocabulaire ? Les révolutions scientifiques, pour reprendre le terme de Kuhn, sont comme les révolutions politiques : elles adorent les nouveaux lexiques et Lyotard (1979) a pu définir les discours scientifiques en tant que «jeux de langage». Or la psychanalyse freudienne montre qu'il est possible de créer un corpus lexical spécifique et autonome — comme le prouve sa vulgarisation — en puisant dans la tradition culturelle et en prêtant attention aux usages communs, la scientificité s'appuyant sur d'autres facteurs.

CONCLUSION : LES RÔLES DE LA MÉMOIRE

Poser la question de la langue de Freud — celle qu'il traite et celle qu'il utilise — ne revêt pas qu'un intérêt linguistique. Un tel questionnement trouve une pertinence immédiate si l'on considère la vague anti-freudienne, largement publicisée, faisant rage actuellement dans les milieux médicaux et universitaires aux États-Unis. L'enjeu est d'ordre thérapeutique : à la méthode psychanalytique, certains courants veulent substituer des approches cliniques «dures» caractérisées par le traitement chimique et l'enfermement et réintégrer dans le strict domaine médical le soin de traiter les troubles psychiques. Mais la donne est également culturelle et socio-politique : un tel tournant signifierait le retour, dans le domaine psychologique, à un savoir médical s'autolégitimant et fonctionnant par son seul discours. Foucault et Deleuze nous ont appris à reconnaître comment, dans les sociétés modernes, la langue est objet de pouvoir pour contrôler les esprits et les individus, avec la complicité des experts et des savoirs disciplinaires (aux deux sens du mot).

³¹ Voir aussi sur ce point H. Meschonnic, *Le langage Heidegger*, Paris, PUF, 1990.

³² Précisons qu'il ne s'agit pas du degré de facilité dans la lecture d'un texte. Adorno n'est pas facile à lire mais la densité de son écriture procède d'un extrême souci de rigueur quant à la menée de ses analyses. On ne saurait reprocher à la philosophie la technicité (et la garantie de sérieux qu'elle suppose) qu'on accorde à d'autres formes de pensée, la réflexion scientifique par exemple. Ce qu'Adorno vise tient dans la distance adoptée par Heidegger quant au monde et au réel.

À l'opposé, la psychanalyse est l'exercice du dialogue et de la rencontre, la croyance dans les bienfaits d'une parole librement énoncée et partagée. Ce que nous avons esquissé sur la charge mémorielle des mots dans la langue freudienne s'éclaire de cette perspective. La mémoire assure plusieurs fonctions. Pour Freud écrivain et penseur, elle est une forme d'intertextualité, qui lui permet à la fois un style propre et une distance par rapport à ce style, le protégeant de la tentation du discours clos et dogmatique. Pour le récepteur et, d'une manière générale, l'individu façonné par le regard freudien, elle inscrit de l'altérité, elle se fait écho, écoute, et se pose d'emblée en interlocution. Si le texte freudien et la pensée analytique développent une langue de mémoire, c'est aussi qu'ils entendent introduire de l'oralité dans l'économie scripturaire et de l'histoire dans l'économie identitaire, assurant au sujet par cette double exigence ce que Freud demandait aux lecteurs dans l'avertissement de *L'interprétation des rêves* : «ne pas refuser au moins à la vie du rêve sa liberté de penser» (1993 : 2).

RÉFÉRENCES

- ADORNO, T. W. (1989) : *Jargon de l'authenticité*, trad. E. Escoubas, Paris, Payot, 199 p.
- BENJAMIN, A. (1989) : *Translation and the Nature of Philosophy*, London and New York, Routledge, 193 p.
- BETTELHEIM, B. (1993) : *Freud et l'âme humaine*, trad. R. Henry, coll. «Pluriel», Paris, Hachette, 207 p.
- CERTEAU, M. de (1990) : *L'invention du quotidien*, t. 1., *Arts de faire*, Paris, Folio/essais.
- DERRIDA, Jacques (1979) : *L'écriture et la différence*, coll. «Points», Paris, Le Seuil, 439 p.
- FREUD, S. (1996) : *Psychopathologie de la vie quotidienne*, trad. S. Jankélévitch, coll. «Petite Bibliothèque Payot», Paris, Payot.
- FREUD, S. (1993) : *L'interprétation des rêves*, trad. I. Meyerson et D. Berger, Paris, PUF, 573 p.
- FREUD, S. (1992) : *Le mot d'esprit et sa relation à l'inconscient*, trad. D. Messier, Paris, Folio/Essais.
- FREUD, S. (1985) : *Trois essais sur la théorie sexuelle*, trad. P. Koeppel, Paris, Folio/essais
- FREUD, S. (1970) : *Ma vie et la psychanalyse* suivi de *Psychanalyse et médecine*, trad. M. Bonaparte, coll. «Idées», Paris, Gallimard.
- FOUCAULT, M. (1994) : «Qu'est-ce qu'un auteur ?», *Dits et écrits*, t. I, Paris, Gallimard, pp. 789-821.
- GOLDSCHMIDT, G.-A. (1988) : *Quand Freud voit la mer*, Paris, Buchet/Chastel, 226 p.
- LEIRIS, M. (1982) : *Mots sans mémoire*, Paris, Gallimard, 154 p.
- LYOTARD, J.-F. (1979) : *La condition post-moderne*, Paris, Éd. de Minuit, 109 p.

- MAHONY, P. J. (1992) : «A Psychoanalytic Translation of Freud», D. G. Ornston Jr. (dir.), *Translating Freud*, New Haven and London, Yale University Press, 251 p.
- MAHONY, P. J. (1990). *Freud l'écrivain*, trad. K. Tran, Paris, Les Belles Lettres.
- MAHONY, P. J. (1982) : «Toward the Understanding of Translation in Psychoanalysis», F. Peraldi (dir.), *Psychanalyse et traduction, Meta*, vol. 27, n° 1, pp. 63-71.
- ORNSTON, D. G. Jr. (1992) : «Bruno Bettelheim's *Freud and Man's Soul*», D. G. Ornston Jr. (dir.), *Translating Freud*, New Haven and London, Yale University Press, 251 p.
- REY, J.-M. (1979) : *Des mots à l'œuvre*, Paris, Aubier-Montaigne.
- RICOEUR, P. (1965) : *De l'interprétation. Essai sur Freud*, Paris, Le Seuil, 534 p.
- SCHNEIDER, M. (1985) : *Voleurs de mots*, Paris, Gallimard.

LES MOTS DE L'OUBLI : LES INTERJECTIONS

Emmanuel HÉRIQUE

Université de Victoria, Canada

1. LES MOTS DE L'OUBLI

1.1 Motivations négatives

Les mots sont sans doute la plus grande entreprise de mémoire de l'humanité : mémoire consciente du texte, rendu accessible par l'invention de l'écrit qui a servi à fixer le récit fondateur (souvent mythe, chronique et loi tout à la fois); mémoire inconsciente de la grammaire que le linguiste cherche à retrouver par un travail d'archéologie du sens et des formes.

Or, dans le panorama impressionnant des études sur le langage il y a une grande absence à l'endroit de ce que l'on nomme «les interjections». Curieux trou de mémoire que nous nous proposons d'explorer ici, en essayant surtout de comprendre pourquoi l'interjection, expression simple, directe, immédiate en apparence, reste toujours très difficile à définir.

Il n'est pas question de faire ici l'historique des études sur l'interjection¹. Un survol de la question nous révèle que les interjections ont été très peu étudiées : quelques monographies, aucun ouvrage décisif, et des articles de teneur très diverse. Pour les grammaires et traités sur la langue, l'interjection n'est pas abordée, ou alors brièvement, et toujours en dernière position. Déjà chez les Grecs, elle était la huitième et dernière partie du langage. Lorsque la question du sens des interjections est abordée, on lit qu'il s'agit de mots «vides de sens», ou alors on décrit les sentiments qu'elles expriment ou la performativité qu'elles opèrent.

La question de la classification des interjections est très ancienne, elle remonte à l'Antiquité. Les divers commentaires, anciens et récents, font remarquer l'origine très diverse des interjections (presque toutes les parties du langage peuvent donner lieu à une

¹On trouvera dans Swiatkowska (1979) une excellente synthèse (I. L'état actuel des recherches sur l'interjection, pp. 49-56). Depuis 1979, on relève divers travaux ayant un certain rapport avec la question des interjections : travaux de pragmatique, recherches sur l'onomatopée, les insultes ou les locutions du dialogue. Pour ce qui est des interjections, on relève notamment : 1) la monographie de Konrad Ehlich sortie en 1986 : *Interjektionen* 2) le numéro de septembre 1992 du *Journal of Pragmatics* consacré entièrement aux interjections (voir Ameka 1992).

interjection). C'est le caractère invariable de l'interjection qui est le plus marquant et qui les rapproche le plus de la catégorie de l'adverbe.

La question du sens des interjections est une question plus moderne. Il est remarquable à cet égard que dans le numéro spécial du *Journal of Pragmatics* (Ameka 1992), les contributeurs posent explicitement la question du sens des interjections, question souvent évitée jusqu'ici².

Résumons : dans l'ensemble les interjections ne sont pas étudiées, et lorsqu'on en parle c'est souvent pour dire que leur sens nous échappe. Dans un tout autre ordre d'idées, nous constatons que les interjections font très rarement l'objet d'une étude systématique dans l'apprentissage d'une langue étrangère, et qu'elles sont presque toujours mal employées par l'étudiant de langue étrangère³. Pour les mêmes raisons, le linguiste qui étudie les interjections ne peut vraiment parler que de celles qu'il pratique le plus authentiquement possible. Ce qui, contrairement aux autres parties du langage, en limite l'étude à celles des langues maternelles.

Il ressort de toutes ces remarques un paradoxe : l'interjection est à la fois la partie du langage la plus intime, la plus «naturelle», et elle est en même temps la partie du langage la moins connue, la moins bien expliquée.

1.2 Motivations positives

L'interjection est unanimement reconnue comme l'expression la plus spontanée des sentiments. C'est d'ailleurs ce qui l'exclut en bonne partie de l'étude et qui la rend si fascinante : elle donne accès au mythe de l'origine des langues, à l'origine «naturelle», animale, du langage.

Il est intéressant à cet égard de relire l'*Encyclopédie* de Diderot. À l'entrée «Interjection», nous lisons cette citation tirée des *Observations sur les langues primitives* de M. le Président des Brosses :

«Les interjections, mêmes telles qu'elles sont dans nos langues formées et articulées, ne s'apprennent pas par la simple audition et par l'intonation d'autrui; mais tout homme les tient de soi même et de son propre sentiment; au moins dans ce qu'elles ont de radical et de significatif, qui est le même par-tout [...]

Elles sont courtes; elles partent du mouvement machinal et tiennent partout à la langue primitive... (p. 935).

² Voir notamment Wierzbicka (1992 : 159-192).

³ On notera à ce sujet la louable entreprise de Jean-Pierre Zerling qui propose une liste d'interjections et d'onomatopées «à l'usage des étudiants étrangers» (Zerling 1995 et 1996). On lit notamment : «Lors de l'enseignement de la phonétique du français il nous a souvent été donné de constater la relative méconnaissance par les étudiants étrangers de certaines catégories de mots. Des mots qui souvent ne font pas partie du vocabulaire enseigné à l'école et qui sont pourtant utilisés quotidiennement par les natifs, pour avoir été appris depuis la plus tendre enfance.» (Zerling 1996 : 158). Cette remarque, faite à propos des onomatopées, est aussi valable pour les interjections.

Le langage d'un enfant, avant qu'il puisse articuler aucun mot, est tout d'interjections...

Le langage des animaux est vraisemblablement tout interjectif» (p. 936).

Dans le même ordre d'idées, voici un extrait tiré de l'article «Interjection» de la *Collier's Encyclopedia* (1968), article par ailleurs long et solidement documenté :

«Being a relic of immemorial ages, when gestures were the only means of expression of man, the interjection may well be the oldest of all parts of speech, and the interjectional theory of origin of language, the so called "pooh-pooh" theory is certainly one that comes closest to the truth» (p. 113).⁴

L'article de l'*Encyclopédie* de Diderot distingue fondamentalement entre «signes naturels» et «signes d'institution», particules «interjectives» et particules «discursives», entre «le langage du coeur» et «le langage de l'esprit».

L'essentiel de la recherche sur les interjections repose sur cette opposition nature/culture, et sur cette fascination de l'origine du parler. D'où les questions fondamentales qui accompagnent souvent la réflexion sur les interjections : l'interjection fait-elle partie du langage ? Comment classifier l'interjection ? Quel sens ont les interjections ? Ces questions ne sont possibles que parce qu'on perçoit une ambivalence fondamentale à propos du phénomène interjectif : comment l'impulsif peut-il s'accommoder d'une construction ? Cette question, qui devient particulièrement pressante quand domine une vision dualiste du monde, trouve un écho plus récent et intéressant dans le livre de Claude Hagège *L'Homme de paroles* :

«Tout serait si simple s'il n'y avait cette tension entre deux extrêmes d'une polarité : le signe motivé, le signe conventionnel ! L'activité remotivante est le produit à la fois d'une tendance régressive de la parole et du besoin expressif de renouveler les formes en les rendant plus solidaires des choses qu'elles représentent, en les faisant habiter par le monde et ses bruits. Ainsi, du conventionnel au conventionnel en passant par le motivé, les langues humaines parcourent indéfiniment une série de cycles [...] Cette oscillation n'a pas de fin. L'homme dialogal a la nostalgie de l'univers [...] Les systèmes phonologiques qu'il construit inconsciemment pour sa langue [...] deviennent pour l'essentiel assez préservés des bruits du monde. Aussi, l'homme fait dominer l'ordre de l'abstraction, construit des taxinomies. Mais il ne renonce pas tout à fait à dire la nature. Sa pratique est rationnelle, mais son instinct le fait parfois incliner à la magie.» (Hagège 1985 : 126-127)

⁴ «Il est tout à fait possible que l'interjection, vestige d'un passé immémorial où l'on ne s'exprimait qu'au moyen de gestes, soit la plus ancienne partie du langage. La théorie de l'origine interjective du langage (la «pooh-pooh theory») se rapproche certainement beaucoup de la vérité». L'expression anglaise «pooh-pooh» est intéressante : verbe fabriqué sur une onomatopée répétée qui exprime le dégoût, le mépris désinvolte et moqueur, on pourrait le traduire par «la théorie du pouah-pouah» ou du «bof», ou même par l'équivalence «la théorie (du) pipi-caca».

Cette longue citation rend compte à la fois du mythe de l'origine motivée des langues, de la tension permanente qui en marque l'évolution, et de la «magie» de la langue qui consiste en signes entièrement (re)motivés. Il s'agit de langue poétique, mythique (Hagège parle du «rêve de la langue-magie»), éminemment subjective et qui poussée à l'extrême tient plus du monologue que du dialogue. C'est dans cette veine que se situe la glossolalie (p. 257), «langue adamique et pré-babélique», «nostalgie d'un paradis perdu». Nous voyons la trace du même rêve dans le succès qu'ont eu les travaux de John Austin et John Searle sur la notion de performativité⁵ : recherche des traces, non plus où le mot colle à la réalité, mais où la réalité colle au mot par la «magie» que lui confère l'institution. Il y a donc de la magie dans les interjections : on croit y entendre le bruissement du monde, et pourtant elles sont entièrement codifiées, puisque différentes d'une langue à l'autre et sujettes à l'invariabilité dans une même langue.

2. LES MOTS DE L'ABSENCE

Sans vouloir prétendre faire ici une étude complète des interjections, nous voulons poser un certain nombre de questions à partir de nos constatations et essayer d'y répondre d'une manière aussi synthétique que possible. Pourquoi les interjections sont-elles doublement un lieu d'amnésie linguistique : amnésie interne, puisque le sens des interjections est dit «vide» ou difficile à préciser; amnésie externe, puisqu'on n'étudie pas les interjections, ou qu'alors elles portent tous les signes du rejet (place finale et réduite dans une étude) ? Par ailleurs, pourquoi cette obsession mythique de l'origine interjective du langage ? Enfin, comment expliquer ce paradoxe que l'interjection exprime la plus forte subjectivité dans la forme la moins flexible ? Amnésie, retour à l'origine, subjectivité sont-ils conciliables en une seule explication ?

Pour tenter de répondre à ces questions, penchons-nous sur certains aspects plus techniques de l'interjection. Si on laisse de côté pour l'instant les considérations récurrentes sur les interjections telles que effets de sens (colère, doute, surprise, etc.) et obsession de l'origine onomatopéique du langage, on peut faire un certain nombre de constatations objectives sur cette partie du langage. Plusieurs faits attirent notre attention :

1) Si l'on se réfère à la nomenclature qui organise l'instance énonciative (le «ego, hic et nunc» de Benveniste), on s'aperçoit que l'interjectif qui se place sur des axes de temps, de lieu ou de personne, privilégie l'absent, le lointain, la troisième personne. En effet, on rencontre ce que nous appelons des «super-interjections» (fréquentes et hautement combinables avec d'autres interjections) comme «ça», «là» et «alors» tandis que «ici» et «maintenant» n'ont pas la valeur abstraite des interjections. Tout au plus pourrait-on rencontrer «ici» ou «maintenant» mis en relief sous forme exclamative afin de renforcer leur sens original non exclamatif. Il y a un degré d'abstraction tout autre avec «ça», «là» ou «alors». On trouve à cet égard d'étonnants commentaires dans *Le Français déchiffré* de Henri Adamczewski (1991) qui présente les langues dans une étude psychomécanique fondée sur la tension thématique/rhématique :

«On aura constaté la très grande richesse d'emplois de *voilà*. Cela n'a rien d'étonnant si l'on se souvient que contrairement à *voici* qui annonce la suite,

⁵ Se reporter aux ouvrages fondateurs : *How to do Things with Words* (Austin 1970) et *Speech Acts* (Searle 1969).

voilà reprend, lie l'énoncé à la situation et au contexte verbal qui précède. Le caractère appréciatif de *voilà* se dévoile dans toute son ampleur, caractère lié tout naturellement à ce qui vient d'être dit à propos de la nature anaphorique de cet opérateur. On comprendra mieux encore le rôle dominant de *voilà* dans le couple *voici/voilà* lorsque le face à face "monde" et "langue" aura été mieux saisi. [...] L'opérateur *là* domine son vis-à-vis *ici* par le nombre d'effets de sens possibles et des emplois de plus en plus métaphoriques.» (p. 61)

Il semblerait que le processus d'éloignement, c'est-à-dire de métaphorisations, soit au coeur de ces interjections à base déictique.

2) Penchons-nous maintenant sur les interjections à base verbale. Il s'agit essentiellement d'impératifs du genre «*va/allez/allons*», «*regarde*», «*écoute*», «*tiens*» où l'on voit un fort processus de métaphorisation. Métaphorisation du contenu lexical (le verbe n'a plus d'objet concret ni dans l'énoncé linguistique, ni dans la réalité extralinguistique); métaphorisation de l'allocutaire qui perd son accord grammatical (cas extrêmes d'absurdités littérales : «*Tiens, voici*»; «*Allez va, ça te passera*»).

3) D'une manière générale, il est reconnu que les interjections sont brèves dans leur forme. Ce phénomène est tout à fait remarquable : il semble indiquer qu'une abstraction du sens (que nous avons appelée jusqu'ici «*métaphorisation*») s'accompagne d'une diminution de la présence physique du mot, d'une espèce d'érosion indicatrice du temps qui passe, et annonciatrice de la «*disparition*» du mot. Le deuxième critère formel sur lequel tous les auteurs sont d'accord, est l'invariabilité de l'interjection. La forme figée est-elle aussi annonciatrice de la mort du mot ?

Les trois points — majeurs, mais non exhaustifs — que nous venons d'examiner tendent à confirmer que les interjections s'organisent autour du sème de l'absent : éloignement de l'instance énonciative, manifeste dans les déictiques (le fait que les déictiques de base soient devenus des interjections est déjà significatif en lui-même) et dans le détournement des rapports allocutifs (cf. impératifs). Absence physique du mot interjectif, qui sort non seulement du rapport allocutif mais aussi du rapport grammatical et syntaxique avec le reste de l'énoncé. Notons qu'à l'inverse, le morphème grammatical est grégaire et agglutinant.

3. LES MOTS DU GROUPE

Cette remarque nous amène à formuler le paradoxe qui entoure les interjections, à savoir que, bien qu'isolées grammaticalement, syntaxiquement et phonétiquement du reste de la phrase mais aussi du reste du système linguistique⁶, elles produisent un effet de sens grégaire. Souvent on rit ou sourit à l'emploi d'une interjection. Alan Gardiner, dans *The Theory of Speech and Language* (1951) définit clairement cette grégarité comme la marque et la condition absolues de l'interjectif. Il écrit, s'opposant avec virulence à Bèhler et Jespersen sur ce point :

⁶ Nous ne nous étendrons pas sur ces points couverts par plusieurs auteurs. Voir notamment pour les questions de morphologie et de syntaxe : Swiatkowska (1979) pp. 64-66. On notera également l'article d'Antoine Culioli (1974) qui fait des remarques de même type «*À propos des énoncés exclamatifs*», dont il analyse les «*schémas circulaires de repérage*».

«Once we are sure that such a cry is intentional, the bridge between the linguistic and the non-linguistic has been crossed [...]

An interjection, as such, does not proclaim or make manifest some single speaker, but all the speakers who have ever used the word. What an interjection names is a specific reaction on the part of any speaker (p. 316)

To qualify as an interjection, a word must needs be habitually, not merely exceptionally, used as an exclamation». (p. 317)⁷

Ces remarques de bon sens sont en étrange contradiction avec les déclarations de l'*Encyclopédie* et les croyances en l'origine spontanée de l'interjection. Elles nous rappellent qu'il s'agit d'éléments totalement construits, formalisés et reconnaissables. Si le mécanisme de l'interjectif exploite l'absent, le locuteur, lui, semble retrouver ses colocuteurs dans une étrange communion.

Nous voulons illustrer ce point par un exemple lié au phénomène interjectif et qui met en évidence à la fois le décrochage allocutif et l'effet de communication — de communion — accrue. Nous voulons parler des jurons. Exclamation à forme interjective et qui exploite les champs sémantiques tabous (religion, sexualité, scatologie), le juron, pour être bien compris, doit se comparer à l'insulte. «Putain» allocuté blessé, oppose les protagonistes; «putain» non allocuté (mais toujours prononcé en présence du même colocuteur) ne doit pas blesser, mais rapprocher les protagonistes dans une «fraternité» qui peut par ailleurs être jugée de mauvais goût. De même pour «con». Notons que le colocuteur peut être masculin ou féminin lors d'une expression comme de l'autre : c'est bien la preuve de la coupure allocutive. Ce mécanisme est à rapprocher de l'usage de la troisième personne en parlant de soi ou de son allocutaire si l'on doit parler d'un sujet trop grave.

C'est donc l'éloignement qui permet le rapprochement, la fiction allocutive qui permet de dire plus. Nous comprenons maintenant la valeur pour l'interjectif de la métaphore dans la situation allocutive.

Il est un champ sémantique intéressant à cet égard : lorsque l'on dit «ce sacré gars», «ce vieux Gérard», «une bouteille de derrière les fagots», c'est clairement le sème de l'éloignement, du caché qui procure plus de plaisir, qui est en jeu. Dans ce contexte, deux adjectifs comme «pourri» et «sale» prennent des polarités différentes : «pourri» peut devenir affectueux, alors que «sale» maintient le rapport allocutif au premier degré. Nous avons parlé de tabou, et les champs sémantiques du juron, nous l'avons vu, sont exclusivement ceux des trois tabous.

⁷ «Dès qu'une exclamation est clairement perçue comme intentionnelle, la ligne est franchie entre le linguistique et le non-linguistique [...] Une interjection, en tant que telle, n'est la marque ni l'expression d'aucun énonciateur en particulier, mais de tous les locuteurs qui l'ont prononcée. Ce qui est nommé par une interjection, c'est une réaction bien déterminée de la part de n'importe quel locuteur. Pour être une interjection, un mot doit être utilisé exclamativement non pas de façon exceptionnelle, mais habituellement.»

4. MORT ET MOTS TABOUS

Ainsi la question qui se pose maintenant est celle du rapport entre tabou et interjection. Tabou, caché, absence, oubli : ce chemin à rebours de notre réflexion montre des liens évidents. Nous pouvons reposer la question : pourquoi cette perte de mémoire à l'endroit des interjections ?

Une relecture de *Totem et Tabou* de Freud (1961) met en évidence certains parallèles frappants : l'activité paradoxale à l'endroit de l'interjectif (subjectivité intense/schémas non allocutés), l'origine mythique et la pureté originelle accordées au langage interjectif, l'absence d'étude rationnelle au sujet des interjections (par exemple le refus d'explorer l'étymologie, comme pour la plus grande partie des interjections), voilà autant d'indices d'une zone linguistique prohibée (ainsi certains mots «sales» ne peuvent pas être dits). Le groupe se retrouve dans le mot-tabou (le juron) même si cette communion n'est pas du goût de tous. Et nous avons vu que précisément le juron n'est que tabou.

Nous arrivons ainsi à une définition de l'interjection et du juron, non plus par leurs effets de sens, mais par les mécanismes spécifiques qu'ils mettent en jeu et par un faisceau de caractéristiques dont beaucoup s'expliquent maintenant :

- forme brève,
- forme invariable,
- métaphorisation,
- fuite du présent de parole.

Il est remarquable que toutes les parties du langage peuvent fournir des interjections — que l'on a appelées pour cela «caméléons». C'est donc dans les opérations que subissent divers «mots» pour devenir interjection qu'il faut chercher une constante.

Il semble que les interjections apparaissent et peuvent disparaître, mais leur forme reste assez stable. C'est le sens qui évolue le plus, devenant par catachrèse⁸ toujours plus rare à saisir. Il semble donc possible de remonter à une source concrète pour la plupart des interjections («Tiens, va, écoute» par exemple). Aussi il existait au Moyen Âge un «tiens» qui accompagnait le don d'un cadeau, puis de quelque chose donné à contrecœur, puis le geste sans l'objet, ensuite l'attitude sans le geste. Attitudes multiples déjà, d'où les nuances possibles. Le sens de surprise, le plus courant aujourd'hui, est un sens moderne, et il est remarquable qu'il s'est accompagné du redoublement du mot, puis d'une répétition multiple («Tiens tiens tiens...»). Peu de mots sont parvenus au stade de la répétition. Faut-il voir une trace interjective dans toute répétition («Ce n'est pas joli joli»; «oui oui») ? Cette question mérite d'être étudiée séparément.

Il semble que pour «Tiens», redoublement et surprise ont évolué de pair. Or le «tiens» de surprise est aussi invariable, c'est-à-dire parfois absurde littéralement («Tiens,

⁸ Voir Darmesteter (1928 : 67-69) : «La catachrèse est l'acte émancipateur du mot; c'est, dans le développement de l'être par germination, la force qui sépare le bourgeon de l'organisme primitif.» Ce langage imagé fait comprendre la force vive qui pousse les interjections, qui n'apparaissent qu'en langue orale, vivante, spontanée et maternelle.

vous ici ?»). Il semble ainsi que le premier terme a été perçu à un moment donné comme mort dans le sens comme il l'était dans la forme, et qu'on a pu bâtir un autre sens sur le premier terme, à l'instar des villes anciennes bâties sur le lieu même d'une autre ville pratiquement ignorée. Cette hypothèse revient à trouver des «morceaux» de langue morte dans une langue vivante — paradoxe curieux, admettons-le, pour ce qui est des interjections.

Il est une autre hypothèse qui apparaît dans certains écrits sur les interjections. Laissons Jacques Derrida, reprenant Jean-Jacques Rousseau, nous la dire :

«Plus une langue est articulée, moins elle est accentuée, plus elle est rationnelle, moins elle est musicale, moins elle perd dès lors à être écrite, mieux elle exprime le besoin. Elle devient nordique». (Derrida 1967 : 344)

Et nous lisons, dans l'étude de Charles-Théodore Gossen sur «L'emploi et la valeur des exclamations et des interjections invocatoires en italien» que «les Méridionaux emploient certainement des invocations plus variées que leurs compatriotes plus sobres du Nord.» (Gossen 1956 : 268-307).

Faut-il penser qu'on s'exclame plus à Tunis qu'à Oslo ? Qu'on interjecte plus à Dakar qu'à Paris ? Il faut se méfier, nous venons de le voir, des impressions que laissent parfois les interjections.

RÉFÉRENCES

- ADAMCZEWSKI, Henri (1991) : *Le Français déchiffré : clé du langage et des langues*, Paris, Armand Colin, 417 p.
- AMEKA, Felix (Ed.) (1992) : *Journal of Pragmatics*, vol. 18, n° 2/3, sept. 1992, «Special Issue on 'Interjections'», XII p. + pp. 101-308.
- AUSTIN, John (1970) : *How to do Things with Words*, Oxford, Clarendon Press, 168 p.
- Collier's Encyclopedia* (1968) : Growell-Collier Educational Corporation, article «Interjections», pp. 113-114.
- CULIOLI, Antoine (1974) : «À propos des énoncés exclamatifs», *Langue Française*, 22, mai 1974.
- DARMESTER, Arsène (1928) : *La Vie des mots*, 17^e éd., Paris, Delagrave, XII + 212 p.
- DERRIDA, Jacques (1967) : *De la grammatologie*, Paris, Minuit, 447 p. (la 2^e partie : «Nature, culture, écriture»).
- DIDEROT et D'ALEMBERT (1778) : *Encyclopédie, ou Dictionnaire raisonné des Sciences, des Arts et des métiers*, nouvelle édition, tome XVIII, Genève, Jean-Léonard Pellet. Article «Interjection», pp. 935-938.
- EHLICH, Konrad (1986) : *Interjektionen*, coll. «Linguistische Arbeiten», n° 111, Tübingen, Max Niemeyer, X + 314 p.

- FREUD, Sigmund (1961) : *Totem and Taboo*, London, Routledge and Kegan, XI + 172 p.
- GARDINER, Alan (1951) : *The Theory of Speech and Language*, Oxford University Press, 2^e éd., XII + 348 p.
- GOSSEN, Charles-Théodore (1956) : «Remarques sur l'emploi et la valeur des exclamations et des interjections invocatoires en italien», *Revue de linguistique Romane*, XX, jan.-juin 1956, pp. 268-308.
- HAGÈGE, Claude (1985) : *L'Homme de paroles*, Paris, Fayard, 314 p.
- SEARLE, John (1969) : *Speech Acts*, Cambridge University Press.
- SWIATKOWSKA, Marcella (1979) : «Quelques considérations sur la nature de l'interjection», *Zeszyty naukowe Uniwersytetu Jagiellonskiego, Prace językoznawcze*, 63, pp. 49-69.
- WIERZBICKA, Anna (1992) : «The Semantics of Interjections», *Journal of Pragmatics*, vol. 18, n° 2/3, sept. 1992, «Special Issue on 'Interjections' edited by Felix Ameka», pp. 159-192.
- ZERLING, Jean-Pierre (1995) : «Onomatopées et interjections en français. Petit lexique phonétique à l'usage des étudiants étrangers (1ère partie)», *Travaux de l'Institut de Phonétique de Strasbourg*, 25, pp. 95-109.
- ZERLING, Jean-Pierre (1996) : «Onomatopées du français : bruits divers et cris d'animaux. Petit lexique phonétique à l'usage des étudiants étrangers (2e partie)», *Travaux de l'Institut de Phonétique de Strasbourg*, 26, pp. 167-170.

LA TERMINOLOGIE MÉTAPHORIQUE D'INTERNET : ASPECTS NOTIONNELS, STRUCTURELS ET TRADUCTIONNELS¹

Ingrid MEYER, Victoria ZALUSKI, Kristen MACKINTOSH et Clara FOZ

Université d'Ottawa, Canada

1. INTRODUCTION

Cet article porte sur la métaphore terminologique, c'est-à-dire celle qui apparaît dans des termes liés à des domaines de spécialité. C'est au domaine d'Internet que nous nous intéresserons afin 1) d'analyser les caractéristiques notionnelles et structurelles des termes métaphoriques produits en anglais dans ce champ, 2) de dégager les stratégies mises en oeuvre pour rendre ces termes en français et 3) d'analyser les facteurs linguistiques et extra-linguistiques qui viennent compliquer ce transfert.

Les recherches entourant la métaphore en terminologie sont récentes, ce procédé, depuis Aristote, ayant plutôt été envisagé comme un procédé littéraire. Ce n'est que depuis les années 1970 qu'il a été traité dans le cadre de la langue générale (Lakoff, 1992) ou, à un degré moindre, dans celui de la langue de spécialité (Ortony, 1979; Lakoff et Johnson, 1980; Liebert, 1994). Plus récemment, lexicologues et terminologues (Assal, 1995; Knowles, 1996; Meyer *et al.*, sous presse; Pavel, 1993; Thoiron, 1994) se sont penchés sur la question sans toutefois toujours l'envisager d'un point de vue interlinguistique, ce que nous tenterons de faire ici.

Pourquoi s'intéresser à la traduction de la métaphore dans un domaine comme celui d'Internet ? Essentiellement, pour contrer le scénario récemment décrit dans *The Economist* (21 décembre 1996) et en vertu duquel les générations futures pourraient n'avoir qu'une seule langue scientifique et technique, l'anglais, les langues «nationales» ne servant plus qu'à la conversation et aux petites choses de la vie quotidienne. Témoin de cette tendance, la pratique, largement répandue dans les domaines de pointe, consistant à employer dans des langues autres que l'anglais des termes forgés dans cette langue. En sorte que pour éviter l'anglicisation tous azimuts de la terminologie, il est essentiel que les langagiers travaillant dans d'autres langues que l'anglais puissent créer des termes adéquats dans leur propre langue. Par «adéquats», il faut bien entendu comprendre des termes dont la désignation soit claire, mais aussi qui «fonctionnent» sur le plan linguistique et extra-linguistique.

Par ailleurs, la métaphore en tant que telle présente un intérêt tout particulier dans le cadre du transfert interlinguistique. D'une part, elle est souvent difficile à traduire (Newmark, 1973), et d'autant plus, nous le verrons, lorsqu'il s'agit d'un emploi spécialisé.

¹ Cette étude a été rendue possible grâce à une subvention de la Fédération canadienne des Sciences humaines et sociales.

D'autre part, il s'agit d'un procédé de plus en plus fréquent dans les domaines de spécialité du fait qu'il permet d'illustrer une notion nouvelle à travers une notion connue et permet ainsi au spécialiste de se rapprocher du lecteur ou de l'utilisateur et donc d'assurer une meilleure communication.

Nous verrons dans un premier temps en quoi consiste la métaphore et pourquoi elle occupe une place aussi importante dans le domaine informatique.

1.1 Qu'est-ce qu'une métaphore?

Fonctions. Du grec *metaphora* («transposition»), la métaphore désigne le procédé consistant à transposer les éléments d'une notion connue vers une notion qui, elle, ne l'est pas. La métaphore a donc une fonction cognitive qui comporte une dimension notionnelle et linguistique. À propos de la première, la philosophie des sciences (Thagard, 1992) nous enseigne que les experts partent intuitivement de la métaphore pour concevoir et développer leurs notions, d'où les recherches actuelles (Liebert, 1994) visant à les aider dans ce sens. Du point de vue linguistique, la métaphore permet de faire «passer» une notion auprès des spécialistes du domaine (ceux des autres champs, l'interdisciplinarité aidant, la récupèrent parfois) ainsi qu'auprès du grand public. La métaphore, enfin, a une fonction esthétique : les juxtapositions auxquelles elle donne lieu étonnent et séduisent tout à la fois. Fascinantes, les métaphores terminologiques d'Internet font d'ailleurs l'objet de commentaires et d'explications dans les revues consacrées au domaine et peuvent donner lieu à toutes sortes de jeux de mots, autre source de difficulté pour qui cherche à les rendre dans d'autres langues.

Types. On peut dire que les termes métaphoriques ont un «cycle de vie» (Ahmad, 1995 : 51). Newmark (1981 : 85) distingue la métaphore *originale* (propre à un sujet et à un contexte donné), la métaphore *récente* (adoptée par d'autres que son auteur) et la métaphore qui est dite *morte* (son caractère métaphorique n'est plus perceptible²). D'où l'importance en traduction de chercher des équivalents lorsque la métaphore est encore récente, c'est-à-dire avant que la forme anglaise ne soit trop ancrée dans une autre langue.

1.2 La métaphore en informatique

Le domaine informatique est particulièrement riche en termes métaphoriques, une richesse attribuable aux fonctions cognitive et esthétique de la métaphore. La première explique que, face à un grand public souvent dépassé par la technique et ses nombreux concepts, les spécialistes aient utilisé la métaphore dans le but d'assurer la compréhension du lecteur et plus particulièrement, dans le cas du vocabulaire propre aux interfaces, de rassurer l'utilisateur : d'où le choix de termes comme *desktop*, *menu* et *wallpaper* pour ne citer que ceux-là. La seconde, la fonction esthétique, a permis de servir une certaine «culture» dite *cyberculture*, marquée par les figures quelque peu iconoclastes de Steve Jobs ou de Bill Gates, et dont la langue évite la scientificité pour une fraîcheur dans l'expression

² Comme le signale Knowles (1996 : 791), les métaphores terminologiques *mortes* peuvent n'être perçues comme telles que par les spécialistes du domaine qui les emploient constamment. À l'extérieur du domaine, c'est-à-dire aux yeux des non spécialistes, leur caractère métaphorique peut persister.

qui va souvent jusqu'au jeu avec les mots : d'où le choix de termes comme *snailmail* et *netsurfing* entre autres.

1.3 Méthodologie, terminologie et structure

Les données que nous analyserons ont été relevées dans un certain nombre d'ouvrages et de magazines anglais et français ainsi que dans des dictionnaires et des glossaires dont certains sont électroniques³. Les données en français proviennent de sources hexagonales et québécoises. Il faut noter par ailleurs que tous les termes analysés n'ont pas forcément pris naissance avec Internet, mais ont pu avoir été empruntés à d'autres domaines (*surf*, par exemple, existait sous la forme de *channel-surfing* dans le domaine de la télévision).

Dans cet article, *terme métaphorique* désigne tout terme contenant des éléments métaphoriques qu'il soit entièrement métaphorique (comme *Web*, *bookmark* et *Gopher hole*) ou partiellement seulement (comme *electronic mail* dans lequel *electronic*, non métaphorique, qualifie *mail*, employé métaphoriquement). Par ailleurs, seules seront analysées les métaphores de type terminologique et non pas celles qui, existant en langue générale, ont été adoptées telles quelles dans le champ d'Internet, et ce même si la distinction n'est pas toujours aisée à établir. Prenons l'exemple du terme *flame war* en anglais : *war*, employé figurativement au sens d'échanges hostiles, n'est pas considéré, dans le cadre de notre analyse, comme métaphorique (encore qu'il faut bien voir qu'il s'agit, dans la langue générale, d'une métaphore *morte*). En revanche, *flame*, dans le sens de «message électronique hostile», est envisagé comme métaphorique, son sens métaphorique étant spécifique au domaine d'Internet.

Dans la deuxième partie de cette communication, nous présenterons sommairement les caractéristiques des termes métaphoriques existant en anglais dans le domaine d'Internet. Dans la troisième partie, nous aborderons les stratégies mises en oeuvre pour les rendre en français et, dans la quatrième partie, nous donnerons un aperçu des facteurs linguistiques et extra-linguistiques qui facilitent ou entravent la création d'équivalents en français.

2. TERMES MÉTAPHORIQUES ANGLAIS DANS INTERNET

La plupart des concepts relatifs à Internet ayant été forgés en anglais, il importe, avant de s'intéresser à leur traduction, de les analyser brièvement d'un point de vue notionnel aussi bien que structurel.

2.1 Caractéristiques notionnelles des termes métaphoriques anglais

Il est clair que dans le domaine d'Internet, comme c'est le cas dans d'autres tels les finances (Knowles, 1996), la bibliothéconomie (Bies, 1996) ou l'intelligence artificielle (Pavel, 1993), les termes métaphoriques sont regroupés autour de diverses thématiques parmi lesquelles nous avons pu relever les suivantes : le **transport** (*to navigate/cruise the Internet*, *Internaut*, *information highway*, *information superhighway*, *to ride/travel the Internet*), le **service postal** (*e-mail*, *snailmail*, *to send/receive/deliver mail*, *mailbox*, *packet*, *envelope*), l'**architecture** (*site*, *under construction*, *gateway*, *bridge*), l'**imprimé**

³ Pour plus de détails concernant ces données voir Meyer *et al.* (à paraître).

(*Web page, Web authoring, bookmark, electronic magazine, carbon copy*), la **communauté** (*virtual community, cybercommunity, electronic neighbourhood*), la **conversation** (*chat, chatroom, talk, netspeak*), le **commerce** (*electronic storefront, electronic mall, shopping cart, e-cash*), le **monde souterrain** ou **interlope** (*network stalking, agent, packet sniffer, cracker*), le **feu** et les **explosifs** (*flame, flamer, mail-bomb, mail exploder*), les **animaux qui creusent** ou **rampent** (*gopher, burrow, spider, ant*).

2.2 Caractéristiques structurelles des termes métaphoriques anglais

2.2.1 Modèles de formation terminologique

Deux catégories structurelles existent selon que tous les composants du terme sont métaphoriques (*termes entièrement métaphoriques*) ou que certaines composantes seulement le sont (*termes partiellement métaphoriques*).

Les termes entièrement métaphoriques peuvent être *simples* ou *juxtaposés*. Dans le premier cas, une seule métaphore est présente sous la forme d'une lexie simple (*Web*) ou composée (*traffic jam*). Dans le second, deux termes métaphoriques sont juxtaposés (*home page*). Pour plus de détails concernant cette typologie voir Meyer *et al.* (sous presse).

2.2.2 Les mots-valises

Comme le montrent les exemples suivants, ce procédé consistant à amalgamer deux mots est fréquent : *brouter* (*bridge + router*), *netizen* (*net + citizen*), *knowbot* (*knowledge + robot*), *internaut* (*internet + astronaut*).

2.2.3 Aspects phonologiques

Nous avons précédemment mentionné l'importance des facteurs esthétiques dans la motivation des métaphores terminologiques d'Internet et à cet égard l'aspect phonologique joue un rôle certain. Pour ne prendre qu'un exemple, citons le cas de *snailmail* dont l'impact tient sans doute autant à la métaphore de l'escargot, symbole de lenteur, qu'à l'allitération contenue dans le terme. Il est clair par ailleurs que c'est par imitation ou analogie que certains termes métaphoriques ont été créés : c'est le cas d'*Internaut* (fondé sur le modèle *astronaut*) ou de *NetPol* (fondé sur *Interpol*). Enfin, l'exemple de *gopher* est doublement intéressant : d'une part la métaphore par laquelle cet outil de recherche (il permet à son utilisateur de remonter une structure hiérarchique pour y repérer des informations) prend le nom d'un petit rongeur vivant dans un terrier est tout à fait parlante. D'autre part, s'ajoute un jeu de mots phonétique entre le nom *gopher* et le verbe *to go for* qui signifie «aller chercher».

3. STRATÉGIES DE FRANCISATION

Les données que nous avons relevées montrent que les stratégies employées pour rendre en français les termes métaphoriques sont au nombre de trois : la première consiste à opter pour une métaphore en français, la deuxième à choisir un équivalent non métaphorique et la troisième à combiner des éléments métaphoriques et non métaphoriques. Étant donné le flottement observable dans le vocabulaire d'Internet, il est

clair que les exemples figurant ci-dessous sont présentés non point comme les seuls équivalents possibles aux métaphores existant en anglais, mais comme des solutions attestées par les ouvrages de référence et repérées dans les sources documentaires⁴.

3.1 Équivalent métaphorique

Quatre cas doivent être envisagés. Dans le premier, la métaphore en français est la même que celle de l'anglais; dans le deuxième, elle diffère de celle-ci; dans le troisième, elle équivaut en partie à la métaphore de départ et dans le quatrième le terme anglais est tout simplement conservé en français.

3.1.1 Choix d'une même métaphore

C'est la stratégie la plus courante, comme en témoignent les exemples de *mailbox* → *boîte à lettres*, *host* → *hôte*, *address* → *adresse* ou *internaut* → *internaute*.

3.1.2 Choix d'une métaphore différente

Ces cas, beaucoup moins fréquents mais intéressants à analyser, sont illustrés par des exemples comme ceux de *gateway* → *passerelle* (deux termes liés à la même thématique et qui désignent des équipements d'interconnexion), *thumbnail image* → *image timbre-poste* (deux termes qui expriment par des images différentes la petitesse) et finalement *spam* → *inonder, arroser* particulièrement intéressant du fait qu'il comporte en anglais une connotation «culturelle» (la marque commerciale de viande en boîte *Spam*) difficilement transposable en français, d'où le choix des équivalents *inonder, arroser* qui rendent bien l'idée d'un système de courrier électronique «inondé» de messages inutiles.

3.1.3 Choix d'une métaphore partiellement équivalente

Dans ces cas, qui relèvent des deux cas précédents, la portée de la notion désignée par la métaphore adoptée en français peut être plus grande ou moins grande que celle du terme métaphorique anglais. Ainsi, dans l'exemple *cruise* → *naviguer, rouler*, le terme anglais s'applique aussi bien au déplacement par voie terrestre que par voie maritime; les termes français s'appliquent en revanche à l'un ou à l'autre, mais non aux deux à la fois. D'où l'emploi de *naviguer* lorsque c'est l'image d'Internet comme «océan de données» qui est retenu, et celui de *rouler* lorsque c'est la notion d'«inforoute» ou d'«autoroute de l'information» qui est plutôt mise de l'avant.

3.1.4 Maintien du terme anglais

La stratégie consistant à conserver en français le terme métaphorique anglais est illustrée par l'exemple de *Web*, généralement maintenu tel quel en français, parfois plaisamment francisé en *ouaibe* ou *ouèbe*. Il faut noter cependant que *toile* s'emploie de plus en plus. *Web* est par ailleurs maintenu dans la forme dérivée *Webmestre* ou *Webmaître* correspondant à l'anglais *Webmaster*. Il faut bien voir cependant que dans les cas où le terme métaphorique anglais est maintenu, il est possible que la métaphore ne soit plus perçue comme telle par un public francophone.

⁴ Pour une liste complète des termes relevés voir Meyer *et al.* (à paraître).

3.2 Équivalent non métaphorique

Il arrive que les métaphores terminologiques soient «perdues» dans le passage de l'anglais vers le français comme en témoignent les exemples de *handshaking* → *entrée en communication* (on ne trouve jamais *poignée de main*), *stale link* → *lien périmé*, *bounce message* → *avis de non-livraison* ou *net stalking* → *harcèlement avec menaces sur réseau* (ce dernier cas ressemblant plus à une paraphrase qu'à un équivalent lexical).

3.3 Combinaison d'éléments métaphoriques et non métaphoriques

Il arrive que les métaphores terminologiques soient maintenues, mais complétées, dans un souci de clarification, par un élément non métaphorique. C'est le cas lorsque le qualificatif *électronique*, *virtuel* ou *cyber* est ajouté aux termes comme dans *signature électronique*, *boîte aux lettres électronique* ou *message électronique*. Autre exemple frappant, celui de *hypertoile*, trouvé comme équivalent de *Web* (plus rarement que *Web*, certes) dans lequel le préfixe *hyper* a été ajouté dans le but de préciser qu'il s'agit d'hypertextes.

4. FACTEURS ENTRAUVANT LA FRANCISATION DES TERMES MÉTAPHORIQUES

Nous verrons maintenant les facteurs qui sous-tendent les stratégies précédemment énoncées.

4.1 Facteurs de type formel

On reconnaît généralement (Newmark, 1988 : 42) que plus les caractéristiques formelles d'un mot ont un impact sur son sens, plus sa traduction peut être problématique. Or nous avons mentionné (section 1) que certaines des métaphores d'Internet devaient leur succès autant à leur forme (esthétique) qu'à leur contenu notionnel. D'où la difficulté de les rendre en français.

4.1.1 Phonologie

Le succès de *snailmail* en anglais (pour lequel il n'existe d'ailleurs aucun synonyme) constitue un exemple parfait de terme métaphorique «réussi» : motivé d'un point de vue notionnel (l'escargot symbolise la lenteur), il doit une grande partie de son succès à son caractère de paronomase, impossible à transposer en français. Un autre exemple comparable est celui du mot-valise *knowbot* formé à partir de *knowledge* et de *robot*. Dans un cas comme dans l'autre les équivalents proposés (*courrier escargot* ou *escargotique* et *logiciel de référence internaute*) sont passablement encombrants et moins «réussis». Il faut cependant noter que pour rendre *email*, *courriel*, terme «réussi», semble de plus en plus utilisé.

4.1.2 Productivité

Dans certains cas, la forme française de base apparaît comme moins productive que la forme anglaise, qu'il s'agisse de dérivation, de préfixes ou de mots-valises.

Dérivation. En anglais, certains termes métaphoriques employés dans le domaine d'Internet ont donné lieu à de nombreux dérivés comme en témoigne l'exemple du nom *flame* qui a donné naissance à *flamage*, *flaming*, *flamer* et au verbe *to flame*. Or si l'on adopte en français la proposition de l'Office de la langue française du Québec⁵ qui suggère de rendre le substantif *flame* par *coup de feu* il paraît difficile de maintenir cet équivalent et d'utiliser des formulations aussi encombrantes que *contenu d'un coup de feu* pour *flamage* ou *acte de lancer des coups de feu* pour *flaming*. D'où les propositions de *flinguer* ou *fusiller* pour *to flame* et de *fusillade* ou *bataille* pour *flame war*. À ce jour, semble-t-il, aucune proposition n'a été faite pour rendre en français *flaming* ou *flamage*.

Préfixes et mots-valises. Dans le vocabulaire d'Internet, sont employés et souvent répétés un petit nombre de préfixes, *cyber*, *net* et *e-* en particulier. Ces derniers donnent parfois lieu à des amalgames (mots-valises). Deux choix sont dès lors possibles en français pour transmettre ces expressions : le calque structurel ou la «re-création». Il arrive aussi que les deux procédés coexistent pour un même terme. On trouve comme exemples de calque avec le préfixe *net*, *netsurfeur* ou *netsurfer*, avec *cyber*, *cyberart* ou *cybercafé* et avec *e-*, *É-zine* (on notera dans ce dernier cas un certain degré de francisation attribuable à l'ajout de l'accent aigu sur le e, francisation également repérable dans le cas de *netiquette*).

Quant aux «re-créations», c'est-à-dire aux dérivations ou aux amalgames spécifiquement français, signalons *inforoutier*, basé sur *inforoute*, *courriel* proposé pour rendre *email* et de plus en plus adopté, *monétique* pour *digicash/ecash/cybercash* ou *internetais* pour *netspeak/cyberspeak*. Certains amalgames, enfin, sont particulièrement délicats à traiter du fait que tant le calque que l'adaptation risqueraient d'entraîner un effet comique non voulu : c'est le cas de *brouter* (combinaison de *bridge* et *router*) pour lequel ni *brouteur* ni encore moins *prouteur* (amalgame de *pont* et *routeur*) n'ont été retenus, la forme *pont-routeur* ou *p-routeur* (ce dernier, à notre avis, est également susceptible de faire l'objet de plaisanteries...) étant plutôt adoptée.

4.1.3 Formes abrégées

De nombreux termes métaphoriques se présentant sous une forme abrégée (parfois acronymique) en anglais, il arrive fréquemment que, faute de pouvoir maintenir en français la métaphore ainsi que l'abréviation, le terme soit repris tel quel comme c'est le cas de *BBS* ou de *POP* (*Post Office Protocol*). Démotivé en français, le terme présente aussi parfois un effet de redondance comme c'est le cas de *Protocole POP*. Certaines abréviations ont par ailleurs été adaptées en français comme c'est le cas de *BBS* pour lequel est utilisé au Québec *BABEL* (acronyme de *BABillard ELectronique*, *babillard* constituant un québécisme pour *tableau d'affichage*). Plus intéressant encore est le cas de *FAQ*, acronyme non métaphorique en anglais pour *Frequently Asked Questions*, maintenu en français, mais correspondant dans cette langue à *Foire aux questions*, véritable «trouvaille». Enfin, des acronymes sont parfois employés en français pour rendre de simples termes anglais : citons à cet égard l'exemple de *BAL* (*boîte aux lettres*) ou celui de *blé* (*boîte aux lettres électronique*) comme équivalent de l'anglais *mailbox*.

⁵ [Http://www.olf.gouv.qc.ca/service/pages/p10cab6.htm](http://www.olf.gouv.qc.ca/service/pages/p10cab6.htm)

4.2 Facteurs de type notionnel

Nous aborderons maintenant les difficultés liées à la *substance* plutôt qu'à la forme. Ces difficultés peuvent être attribuables à l'anisomorphisme des systèmes linguistiques aussi bien qu'à des facteurs culturels ou à la question de l'étendue des métaphores.

4.2.1 Différence d'extension

Étant donné l'anisomorphisme des systèmes linguistiques, il arrive qu'une métaphore en anglais n'ait pas d'équivalent en français. Aux exemples de métaphores partiellement équivalentes présentées à la section 3.1.3 (*naviguer/rouler* pour *cruise* ou *fureter* pour *browse*) peuvent s'ajouter les exemples de *album électronique* pour *electronic coffee table book* ou de *binette/émoticon* pour *smiley*. Dans ce dernier cas, toutefois, il faut bien voir que si l'on traduit *smiley* par *souriant* (ce qui est parfois le cas), la question de l'anisomorphisme des systèmes linguistiques ne se pose plus. Parfois, le problème est insurmontable comme en témoigne l'exemple de *cobwebsite*, métaphore qui désigne en anglais un site web non mis à jour et qui se fonde sur la différence existant dans cette langue entre les substantifs *cobweb* (toile d'araignée poussiéreuse et ancienne, caractéristique des maisons hantées) et *spiderweb* (toile d'araignée plus récemment tissée). Cette différence n'existant pas en français (une seule expression existe, *toile d'araignée*), impossible de trouver dans cette langue un véritable équivalent.

4.2.2 Facteurs culturels

Les différences culturelles constituent bien souvent des entraves à la traduction et l'un des exemples les plus intéressants est celui de *spam* (et de son dérivé *spamming*) qui est l'équivalent, dans le domaine informatique, de «junk mail» et désigne donc une surabondance de courrier électronique qui dérange celui qui le reçoit. Aux États-Unis comme en Grande-Bretagne le mot évoque une marque de viande en boîte qui date de l'époque de la seconde guerre mondiale au cours de laquelle la viande était rare, d'où la création d'un produit en conserve. Plus récemment, le mot *spam* a été repris par les Monty Python, ce qui a contribué à familiariser la jeune génération à la notion et à son expression. Dans le monde francophone, en revanche, il est clair que le terme n'a pas de résonance, d'où les «traductions» *inondation*, *arrosage* ou *multi-postage excessif*.

4.2.3 Métaphores filées

Il arrive qu'une métaphore particulièrement réussie soit «filée», c'est-à-dire s'étende sur plusieurs mots dans un même champ associatif. Le terme *gopher*, par exemple, désigne un outil de recherche ainsi baptisé du fait qu'il a été inventé à l'Université du Minnesota, dont la mascotte est le petit animal nommé en anglais *gopher* (*spermophile* en français). De plus, cet outil permettant de «creuser» dans Internet à la recherche de données, le terme de *gopher* désignant un petit animal vivant dans un terrier était particulièrement adapté en anglais. En français, cependant, le terme a été adopté comme un simple nom propre (ce qu'il est aussi en anglais) sans que soit perçu son caractère métaphorique. Quant aux autres créations fondées sur *gopher* (*gopher hole*, *gopherspace*, *to tunnel/burrow*) elles sont généralement maintenues en français et accompagnées d'une explication pour les deux premières et rendues par *fouiller* ou *creuser* pour les deux dernières.

4.3 Facteurs régionaux

Notre étude se limite aux usages français et québécois, mais ne serait-ce qu'entre ces deux variétés de français certaines différences peuvent être observées : nous avons déjà mentionné l'exemple de *BBS*, maintenu tel quel en français hexagonal et rendu par l'acronyme *BABEL* au Québec. Cet exemple illustre peut-être la tendance consistant au Québec à éviter l'anglicisme, contrairement à une tendance française qui veut que l'on s'en accommode fort bien. L'exemple du verbe *to surf* est à cet égard significatif : bien que le verbe *surfer* soit attesté dans les grands dictionnaires de langue française, il semble que les auteurs québécois soient plus réticents que leurs confrères français à employer des expressions comme *surfer sur le Web* ou *netsurfeur*.

5. DISCUSSION

Si la métaphore occupe une place considérable dans la terminologie du domaine d'Internet c'est qu'elle permet de faire «passer» de nouveaux concepts (fonction cognitive) tout en évitant la trop grande technicité. Il faut bien voir par ailleurs que dans le domaine d'Internet s'ajoute une dimension ludique qui donne lieu à des jeux de mots ainsi qu'à des «trouvailles» linguistiques. Nous avons vu que le passage de l'anglais vers le français de ces termes métaphoriques était soumis à un certain nombre de contraintes de type notionnel, linguistique ou culturel, mais qu'en général on tend à proposer en langue d'arrivée un terme également métaphorique, y compris d'ailleurs dans des cas où l'anglais ne contient pas de métaphore (voir l'exemple de *FAQ/FAQ*). Sans doute cela est en partie dû aussi au fait que la métaphorisation constitue, avec la spécialisation sémantique, un des mécanismes de formation de la langue technique française (Quemada, 1978 : 1166-1176)

Une autre question se pose, celle de l'incidence des métaphores sur d'autres parties du discours : on constate par exemple en anglais qu'avec le substantif *Web* l'usage dans le champ d'Internet veut que l'on dise «*on the Web*» alors que la préposition normalement employée avec ce substantif est «*in*» (comme dans «*a fly was caught in the spider's web*» ou dans «*caught in the web of words*»). De même, en français, la pertinence de la préposition *sur* employée avec les termes *Web*, *navigation* ou *naviguer* fait l'objet de discussions⁶.

Les données que nous avons pu recueillir pour notre étude (limitée, certes) permettent de penser que dans le domaine d'Internet les possibilités expressives de la langue française en matière de terminologie métaphorique sont d'ores et déjà bien exploitées. Il nous reste à espérer que d'autres études verront le jour dans le domaine en constante évolution que constitue Internet et qu'elles viendront confirmer cette tendance.

⁶ Voir en particulier les chroniques de l'Office de la langue française du Québec à l'adresse <http://www.olf.qc.ca/service/pages/p10cab13.htm>

RÉFÉRENCES

- AHMAD, Khurshid (1995) : «Pragmatics of Specialist Terms : The Acquisition and Representation of Terminology», Steffens (Ed.), *Machine Translation and the Lexicon*, Proceedings of the 3rd International EAMT Workshop, Heidelberg, 1993, Berlin, pp. 51-76.
- ASSAL, Allal (1995) : «La métaphorisation terminologique», *Actualité terminologique*, 28 (2), pp. 22-24.
- BIES, Werner (1996) : «Thinking with the Help of Images: On the Metaphors of Knowledge Organization», *Knowledge Organization*, 23 (1), pp. 3-8.
- KNOWLES, Frank (1996) : «Lexicographical Aspects of Health Metaphor in Financial Text», Gellerstam *et al.* (Eds), *EURALEX '96 Proceedings I-II*, Papers submitted to the Seventh EURALEX International Congress on Lexicography, Göteborg, Sweden, Part II, Göteborg, Göteborg University Department of Swedish, pp. 789-796.
- LAKOFF, George (1992) : «Metaphor and Semantics», Bright (Ed.), *International Encyclopedia of Linguistics*, vol. 2, Oxford, Oxford University Press, pp. 417-419.
- LAKOFF, George et M. JOHNSON (1980) : *Metaphors We Live By*, Chicago, University of Chicago Press.
- LIEBERT, Wolf-Andreas (1994) : «Lascaux—A Hypermedia Lexicon on Metaphor Models for Scientific Imagination», Martin *et al.* (Eds), *EURALEX '94 Proceedings*, Paper read at the 6th EURALEX International Congress, Vrije Universiteit, Amsterdam, September 1994, Amsterdam, Vrije Universiteit, pp. 494-500.
- MEYER, Ingrid, ZALUSKI, V. et K. MACKINTOSH (sous presse) : «Metaphorical Internet Terms: A Conceptual and Structural Analysis», *Terminology*.
- MEYER, Ingrid, ZALUSKI, V., MACKINTOSH, K. et C. FOZ (à paraître) : «Metaphorical Internet Terms in English and French».
- NEWMARK, Peter (1973) : «An Approach to Translation», *Babel*, vol. 19, n° 1, pp. 3-19.
- NEWMARK, Peter (1981) : *Approaches to Translation*, Oxford, Pergamon Press.
- NEWMARK, Peter (1988) : *A Textbook of Translation*, New York/London, Prentice Hall.
- ORTONY, Andrew (Ed.) (1979) : *Metaphor and Thought*, Cambridge, Cambridge University Press.
- QUEMADA, Bernard (1978) : «Technique et langage», *Histoire des techniques*, Encyclopédie de la Pléiade, Paris, Gallimard, pp. 1146-1240.
- PAVEL, Silvia (1993) : «Neology and Praseology as Terminology-in-the-Making», Sonneveld et Loening (Eds), *Terminology: Applications in Interdisciplinary Communication*, Amsterdam/Philadelphia, Johns Benjamins, pp. 21-34.
- THAGARD, Paul (1992) : *Conceptual Revolutions*, Princeton, N.J., Princeton University Press.
- The Economist* (1997) : «The Coming Global Tongue», pp. 75-77.
- THOIRON, Philippe (1994) : «La terminologie multilingue : une aide à la maîtrise des concepts», *Meta*, 39 (4), Montréal, Presses de l'Université de Montréal, pp. 765-773.
- ZALUSKI, Victoria (1997) : *Metaphor-based Internet Terms in English and in French*. M.A. thesis, School of Translation and Interpretation, University of Ottawa, Canada.

APPROCHE CULTURELLE DE LA DÉNOMINATION EN TERMINOLOGIE

Atibakwa B. EDEMA

CELTA, Kinshasa, République Démocratique du Congo et LLACAN/CNRS, Paris, France

INTRODUCTION

«La leçon des mots, quand on les chatouille un peu, est aussi leçon d'histoire.»
(A. Rey, 1996 : 13)

Quand des missionnaires européens montrèrent au roi des Kuba (sud-est du Congo) une *motocyclette* neuve qu'ils venaient tout juste d'acquérir, «ce nouvel objet de la technologie moderne n'attira guère l'attention du roi. Ce qui le fascinait bien plus, c'était les traces laissées par les roues de la moto sur le sable [...] et il en recopia aussitôt les dessins¹»; dessins qui s'ajoutèrent aux nombreuses variétés de fresques qui faisaient déjà la réputation de ce peuple d'artistes depuis des générations.

Chez les Liko (nord-est du Congo), le *vélocipède*, ancêtre de la motocyclette (ou *bicyclette motorisée*), reçut le nom de **magú-mákwanganya**, mot à mot, «jambes quatre», ce qui peut mieux se traduire en bon français par *quadrupède*.

En Centrafrique ce nouveau véhicule fut baptisé, en sängö, au nom de **gbâzâbândâ**, littéralement «roue(s)-caoutchouc», ou *roues en caoutchouc*.

Au Mali, la nouvelle «bête» fut nommée **nàgàsó**, *cheval de fer*, par les locuteurs bambaraphones.

En wolof (Sénégal), bien que ce soit le mot d'origine française qui désigne la bicyclette, **weló**, il existe un proverbe qui dit «acheter un *cheval en tige* en attendant d'acheter un *cheval de fer*²».

Avec ces quelques exemples, on voit comment un *même objet*, entré presque au *même moment*, a pu recevoir *différentes dénominations*.

¹ S. Curtil (1991) : *N'tchak, un pague de fête au pays des Kuba*, Paris, Collection Kitadi, Musée Dapper.

² Nous remercions J.L. Diouf, linguiste sénégalais, de nous avoir donné cette information.

OBJET

Quand un nouveau produit entre dans un univers dans lequel il n'a pas été conçu ni fabriqué, trois possibilités au moins (cf. *infra*) peuvent se présenter pour son «baptême» :

- soit que le produit entre avec son nom d'origine, moyennant ou non une adaptation morphophonologique du nom emprunté; ainsi l'appareil *radio récepteur* a-t-il (presque) gardé son nom **ladió** en lingala, **aladiyó** en ɔtɛtɛ́la au Congo (Caprile et al. 1979);

- soit que le nom d'origine est «traduit» ou «adapté» dans la langue d'adoption; *avion* est «traduit» en lingala par **nedek** (même référent, «oiseau»); en swahili un étudiant a fait une traduction adaptée de *géométrie* par **elimu ya mistari** ou *science, discipline des lignes* (CELTA, 1974 : 23);

- soit, enfin, que les locuteurs recourent à la créativité lexicale «interne»; le cas de noms de plantes vivrières «importées» d'Amérique latine est particulièrement éloquent en Afrique³ (Edema et Epanga); **nsɔngɔ** ou **monsɔngɔ** pour *manioc* en lingala; **mpunga** pour le *paddy* en kiswahili, pour ne citer que ces quelques exemples.

Il y a une quatrième possibilité dont nous parlerons plus loin (cf. *infra* § sur *l'ancien et le nouveau du même*).

La présente communication a justement pour ambition de présenter quelques types de dénomination en terminologie, en exploitant deux processus d'évolution lexicosémantique, la **démotivation** et la **remotivation**. Nous nous baserons sur un point de vue culturel, en s'écartant le plus possible des deux premières possibilités présentées ci-haut, à savoir l'emprunt et la traduction. Cette étude n'est donc qu'un prolongement de celle qui a déjà fait l'objet de réflexion par Diki-Kidiri (1996).

En prenant appui sur des exemples de quelques langues africaines différentes, nous essayerons de voir comment, dès l'instant où le culturel, mêlé à l'expérience, entre en terminologie, s'est réalisée la dénomination dans ces langues, plus précisément comment à été appréhendé le nouveau en partant du passé. Le but ultime est de voir, parmi les différentes possibilités qui s'offrent au terminologue (emprunt, traduction), en sa qualité de porteur de double casquette, quand il veut dénommer de nouvelles notions ou de nouvelles fonctions, celle qui relève soit d'une récupération des mots vieillis, soit d'une mutation des sèmes des mots actuels.

MÉTHODOLOGIE

Dans Caprile et al. (1979), les auteurs avaient préconisé «la création lexicale spontanée en Afrique centrale par emprunt au français⁴». Découlant du projet initié par

³ En mai 1994 a été organisée une table ronde sur ce thème intéressant intitulé «Plantes, paysages et histoire en Afrique sub-saharienne»; malheureusement les actes de ce colloque n'ont par encore été publiés.

⁴ Il y a quelque chose comme une réunion des contraires dans l'intitulé même de ce projet : «création par emprunt...». Quand on crée on n'emprunte pas, on invente.

l'ACCT sur les «lexiques thématiques de l'Afrique centrale», ces travaux se sont consacrés à «l'adaptation lexicale des langues africaines à certaines réalités de l'évolution, aux conséquences de nouvelles fonctions assumées par les langues africaines, aux problèmes de traduction des mots étrangers sur le plan technologique» (Caprile, 1979 : 9).

Des lexiques ont été ainsi publiés par la suite, en tenant compte cette fois de la création lexicale «interne» et de la néologie. Certaines propositions de traduction, d'adaptation, de création interne et de néologie lexicale étaient plus ou moins heureuses. Dans l'ensemble, ces travaux ont été faits sans méthode précise ou alors en faisant de l'emprunt, de la traduction ou de l'adaptation des principes de travail. Ainsi, en swahili par exemple, dans le domaine des *activités économiques et sociales*, l'item proposé pour dénommer la notion de *commerce extérieur* a été traduit mot à mot en **biashara ya inje** (*commerce de l'extérieur*). Une traduction comme **biashara na ugeni** (*commerce avec l'étranger*) eut été plus appropriée.

La présente analyse se propose, elle, d'explorer non pas seulement la spécification lexico-sémantique autour d'un champ lexical que peuvent exploiter les langues de spécialité (médecine, linguistique, informatique, sport, art...) mais de puiser dans les **réserves linguistiques, culturelles, socio-sémantiques et historiques que les locuteurs (ou la langue) gardent en mémoire.**

En effet, l'exercice auquel s'étaient livré des étudiants de Lubumbashi en traduisant dans des langues locales des notions apparemment éloignées des besoins immédiats des populations congolaises, comme la *physique nucléaire*, le *marxisme* et l'*existentialisme*, etc. (CELTA, 1975 : 23-62) ou l'ensemble de traductions en wolof dans plusieurs domaines scientifiques proposées par Diop (1975 : 154-233) débordent largement du linguistique dont il relève d'abord comme fait du langage.

Cet exercice gagne à être approfondi.

C'est pourquoi depuis Lyon (Diki-Kidiri et al. 1997), nous avons voulu nous éloigner le plus possible de l'emprunt et de la traduction pure et simple, en recourant à ce que nous appelons maintenant «démarche culturelle de la dénomination».

Sans les écarter *a priori*, notre démarche nous semble différente de la **polysémisation** pure et simple, sinon elle serait simplement **synchronique**, ou de la **paronymie** qui font naître les mots dans un même champ lexico-sémantique, sinon elle serait de la **dérivation** uniquement.

En effet, les exemples polysémiques recourant au même signifiant (par métonymie) du type du lingala **mbúla** signifiant à la fois *pluie* et *an, année* ou en liko **syángásu** signifiant *saison sèche* et *an, année* ne nous semblent pas très opératoires ici, parce qu'ils concernent une évolution d'origine interne de la langue. En second lieu, on ne devrait donc recourir à la dérivation que si elle s'accompagne des recherches philologiques.

La préférence méthodologique est donc la suivante :

1° **motivation philologique**, ce qui pose certes un problème de taille, celui des sources dans les sociétés à tradition orale comme les nôtres;

- 2° **métaphorisation**; langage, à coup sûr, motivé;
- 3° **motivation morpholexicale** (dérivation et composition).

DÉMARCHE CULTURELLE

Au delà des propositions plus ou moins heureuses qui ont été faites pour faire appréhender la notion d'informatique à un utilisateur lambda dans sa langue maternelle (Diki-Kidiri et al. 1997), il s'est dégagé peu à peu une démarche que nous disons culturelle et que nous ne pouvons mieux faire comprendre qu'au travers des exemples tirés de langues africaines, en recourant, le cas échéant, au français.

Mais, auparavant, posons d'abord le principe central de la démarche; il s'énonce comme suit : **la reconceptualisation opératoire de l'objet à dénommer compte plus que la dénomination elle-même**, c'est-à-dire plus que le nom qu'on lui attribue. En fait, la reconceptualisation de l'objet, **aidée ou guidée par une culture interne, précède la dénomination**.

Une autre condition de la démarche culturelle est qu'ici l'objet doit être, à l'origine, **extérieur** à la culture du peuple qui l'adopte et qui prouve sa domestication en le dénommant par le recours à sa propre culture. Dans ce cas, la culture est le principal instrument d'appropriation du concept ou de l'objet, la langue ne faisant que l'actualiser.

Tout cela se traduirait par l'attitude suivante : **si j'ai bien compris (saisi) le fonctionnement de cet objet, nouveau pour moi, si j'ai bien redessiné (déconstruit) le référent, si je l'ai bien concrétisé dans mon esprit, en me servant de ce que je possède (= culture), ce nouvel objet que je vois aspire à être dénommé comme tel**. Certes, il ne s'agit pas d'une déconstruction **complète** de l'objet mais de ses **parties essentielles** (en l'occurrence, seule la déconstruction du vélo par les locuteurs bambaraphones nous semble complète).

Pour que le choix dénommatif proposé ait des chances d'être facilement compris et surtout soit adopté, il faut donc que le signifiant soit adapté aux réalités culturelles du locuteur. Cela exige des connaissances précises sur les modes d'appréhension et de représentation de différentes cultures, de l'histoire du peuple et de la langue dans sa profondeur de temps.

D'où l'intérêt de partir de l'intérieur de la langue, de son bagage socioculturel ou plus précisément des bases culturelles de la reconceptualisation (récognition) des objets oubliés en vue de les intégrer dans le schéma des objets actuels. Ce qui impliquerait, par exemple, une résurrection de la technologie ancienne (endogène) même si elle ne s'accorde pas totalement avec une technologie importée, toutes choses n'étant pas toujours égales par ailleurs.

Par sa conception, il s'agit d'une démarche associant à la fois le linguistique, les données culturelles et historiques de chaque peuple. En effet, un locuteur sahélien aura une autre conception d'un objet qu'un locuteur de l'Afrique centrale devant le même objet. Nous l'avons vu avec l'exemple de *vélo* ci-dessus.

Dans l'esprit de notre démarche, la dénomination par remotivation / démotivation devrait donc se faire principalement par **la réactivation des mots déjà formés mais dont l'usage a été oublié**; en second lieu par **métaphorisation des objets culturels de conception interne**; en troisième lieu par la dérivation et la composition.

Ainsi les lexies du lingála telles que, **nganga** *guérisseur traditionnel, féticheur, devin* et **mónganga** *médecin formé à l'école européenne*, apparaissent comme des vestiges, par démotivation et remotivation, du mot reconstruit **monganga** dont le sens initial serait *initié*.

La discussion à caractère historique concernant les tons des préfixes nominaux de la classe 1 en lingála ne sera pas approfondie ici. Les habitués de langues à tons auront cependant remarqué la différence tonale entre la lexie actuelle **mónganga** et celle que nous avons reconstruite **monganga**. Sauf dans de rares cas, d'ailleurs de création récente, les préfixes nominaux de la classe 1 sont de tons bas en lingála, tout comme en proto-bantu d'ailleurs. On peut seulement regretter que le proto-bantu soit plus reconstruit morphologiquement et très peu sémantiquement. On pourrait en effet aboutir à ce que nous avons appelé «parasynonymie horizontale» c'est-à-dire «phénomène de "réflexe lexical" dans deux langues d'une même famille» (Edema et Nduku 1993 : 355-372). Ainsi entre le lingála et le lilikó, en partant des recherches onomasiologiques uniquement, on ne saurait trouver un même étymon pour *cheveux* car pour le lingála c'est **súki** et pour le lilikó c'est **túkату**. Seules des recherches sémasiologiques devraient rapprocher ces deux langues car en lilikó **sóki** désigne les poils du pubis. De même, entre le swahili et le lingála, si on recherchait un mot commun pour *sein* (ou lait) on aurait **mabele** pour le lingála et **maziwa** pour le kiswahili. Mais en faisant appel à un vaste champ à la fois sémantique et cognitif on découvrira que **mbele** du kiswahili, *avant, devant*, se rapproche de **mabéle** du lingála, sachant par ailleurs que certaines parties du corps font office de prépositions ou de directionnels dans certaines langues africaines. La ressemblance morphologique fortuite est donc à écarter au profit d'une origine commune.

Le concours du comparatisme et de la grammaire historique sera donc d'un grand apport dans la dénomination. Par exemple, les affixes en bantu n'ont pas encore bénéficié d'une attention aussi minutieuse qu'ils le méritent. À l'instar de l'étude sur les adverbes en *-ment* en français, une étude faite par Kamba (1981) sur les morphèmes de négation dans certaines langues bantu donne pourtant une indication pour approfondir la mémoire des mots. En effet, selon Kamba les morphèmes *-ta-*, *-me-*, *-to-* indiquant respectivement le futur, le perfectif et la négation en swahili viendraient des verbes **kutaka** *vouloir*, **kumaliza** *terminer* et **kutoa**, *trahir*. À l'exception du dernier, les autres verbes sont tous des auxiliaires.

L'ANCIEN ET LE NOUVEAU DU MÊME

En fait, à part les trois possibilités que nous avons présentées *supra* (emprunt, traduction et création interne) qu'arrive-t-il quand une réalité, apparemment importée, n'est pas si nouvelle que cela ?

Quand on ouvre un dictionnaire actuel du lingála, dans la partie lingála-français, à l'article **nganga** (**nkísi**), on lit les définitions suivantes : *devin, médium, féticheur* avec

ce commentaire : «supposé guérir au moyen d'incantation». Par contre, dans la partie français-lingala, à l'article *médecin*, c'est le mot **mónganga** qui lui correspond. Il faut noter que l'auteur ne mentionne aucun renvoi à l'une ou l'autre lexie (genre «dérive de») pourtant morphologiquement si proches.

Quelle explication donner à cette lacune ?

En fait quand on lit l'histoire on remarquera que dans un premier temps (époque pré-coloniale) quatre spécialités : **pharmacopée, médecine, divination, prêtrise** étaient concentrées aux mains d'une même personne appelée **nganga-nkísi**. Puis dans un second temps, à l'époque coloniale, il s'est produit un dédoublement de la fonction par l'apparition des prêtres européens, qui sont des anti-féticheurs, ou plus précisément qui ne sont pas des médecins. Ce dédoublement a conduit à :

- 1° la création lexicale de **nganga-nzámbe** (= médecin de Dieu) par opposition à **nganga-nkísi**;
- 2° la polysémie du mot **sángó**, *père*, jusqu'à la perte de son sens de *père géniteur* pour ne désigner plus que le *prêtre* (affilié à *père*);
- 3° la paronymie (parasynonymique) de **mónganga** (création récente);
- 4° l'opposition **mónganga** vs **ngangá (nkísi)**, tous deux, à n'en pas douter, médecins, même si le deuxième y ajoute des qualifications que le premier n'a pas.

En réalité, ce qui est en cause dans cette opposition, c'est la **médecine moderne** face à la **médecine traditionnelle**, le **prêtre** face au **guérisseur**, le **nouveau** face au **moderne**, le **passé** face au **présent**, le **modernisme** face à la **tradition**, la religion **catholique** face à ce qu'on appelle l'**animisme**.

Un autre exemple : quand on consulte un vieux dictionnaire du lingala comme celui de Courbon, publié en 1908, dans la partie lexicale, le verbe *écrire* est rendu par l'équivalent lingala **kusona**. On ne sait pas depuis quand **kokoma** est venu se substituer à **kusona** et ce qu'il signifiait avant, ce qui serait très important.

Pour *lire*, c'est le verbe **kutunga** qui est mentionné alors qu'actuellement c'est **kotángá**, qui signifie par ailleurs *compter*, qui est maintenant employé.

Chez les Liko par còtre, *écrire* a été emprunté du swahili **kuandika**. Pour *calligraphier*, «écrire avec soin», ils se sont servis de leur arsenal linguistique propre, en utilisant le verbe **kóing**, qui signifiait se maquiller⁵ à l'aide d'un stylet. Ce genre de maquillage, jugé rétrograde, a été totalement abandonné. Il a été repris pour dénommer une opération, pour nouvelle qu'elle soit, ne faisait pas moins appel à une pratique ancienne.

LA DÉNOMINATION

Mais avant de dénommer, quels sont les préalables de l'opération ?

La dénomination se fait après l'opération déconstruction-reconstruction, opération que Diki-Kidiki nomme «reconceptualisation» (ici même).

⁵ On notera au passage la démotivation entre «maquiller» et «se maquiller» en français.

En verbalisant de nouveaux objets, réels ou épistémologiques, l'homme élargit ses horizons cognitifs et, de ce fait, son vocabulaire dont le volume suit la courbe de ses découvertes. Tout nouvel objet découvert ou conçu doit se «vêtir» d'un signifiant, qu'il faut créer dans un premier temps. Plus tard, soit par analogie, soit plutôt parce que le mot recèle encore des parcelles sémantiques inexploitées, un même signifiant endossera plus d'un signifié, ce qui explique la polysémie. Le mot, tel une fleur, dévoilera, au fil du temps et des occasions, ses divers pétales sémantiques, ceux-ci pouvant se perdre dans la profondeur des temps. Seul l'étymologiste peut ressusciter les mots ou les sens perdus, à l'exemple de «ordinateur» ci-dessus (Marcellesi, 1979) où le terminologue modifie ceux qui sont actuels (cf. par exemple le vocabulaire de l'informatique, *fichier*, *lecteur*, *serveur*, etc.). La langue s'enrichit ainsi au fur et à mesure que les besoins expressifs des locuteurs s'élargissent et que de nouveaux objets entrent dans leur univers.

Qu'est-ce alors que la dénomination ?

De façon générale, la dénomination concerne trois domaines spécifiques : la philosophie, la génétique et la **linguistique**.

Dans *Encyclopédie philosophique universelle*⁶ la «dénomination» est définie comme un «acte par lequel on donne un nom à quelque chose»; c'est aussi le «résultat de cet acte, c'est-à-dire nom donné à quelque chose. Comme relation, la dénomination va de la chose au nom ; elle est donc l'inverse de la dénotation».

De façon générale, la dénomination n'est qu'une application de l'ononastique sur les objets plus ou moins matériels. Elle est donc liée à une des fonctions du langage qui est de nommer les choses du monde, après les avoir identifiées, c'est-à-dire distinguées de celles déjà connues, bref après les avoir reconceptualisées. La dénomination relève donc du cognitif. Foucault (1990) la nomme *mathesis* c'est-à-dire la «science de la mesure et de l'ordre», par le principe de «l'identité et de la différence» entre les choses.

En linguistique la dénomination se définit comme le «processus par lequel les mots (qui malgré l'étymologie, ne sont pas toujours des noms) sont affectés aux référents. Par extension, *dénomination* s'applique parfois au mot lui-même dans la mesure où il vise le référent» (Arrivé et al. 1986 : 213).

Échappant au principe de l'arbitraire, établi par Saussure, la dénomination rend nécessaire le signe linguistique, selon l'affinement du principe apporté par Benveniste (1966 : 49-55). Cette correction, rend complexe le signe linguistique puisqu'elle introduit une troisième dimension comme nous le démontre Diki-Kidiri ici même. De par son fonctionnement, le signe dénominatif, efface donc plus ou moins «l'arbitraire du signe» puisqu'il n'est plus «gratuit».

Considéré de ce point de vue, le signe linguistique n'est plus biface, tout au moins dans les relations dénominatives.

Cependant si «la dénomination des objets paraît à première vue une tâche simple se prêtant à une description simple», en imaginant «que dénommer un objet consisterait

⁶ Paris, PUF, Tome 1 *Les notions philosophiques*, 1990, pp. 591-592.

d'abord à le reconnaître, puis à lui trouver un nom» (Morton, 1984 : 19), «en linguistique, la dénomination est un concept aux contours mal délimités dont l'extension varie considérablement selon les théories et les auteurs. Les définitions "larges" la présentent comme la relation qui unit une expression linguistique à une entité extra-linguistique; les définitions "moyennes" l'assimilent au rapport qui s'établit entre unité codée, item lexical en tête, et son référent; les définitions "restreintes", enfin, la limitent au lien désignationnel entre la catégorie grammaticale nominale, dans laquelle on privilégie le substantif, et la classe ou catégorie référentielle correspondante» (Kleiber, 1984 : 77). Ce qui aboutit, conclut, Kleiber, à la «circularité» entre «nom» et «dénomination», entre le résultat et l'acte.

Aussi, pour éviter les divergences de ces différents types de définitions, retient-on souvent le «caractère nécessairement **motivé** ou non du nom attribué» (par ses qualités ou par sa fonction). Définie en linguistique comme «relation d'un signe entre sa forme et son contenu», la **motivation** est en effet, la base de la dénomination (par dérivation et par la composition) parce qu'elle permet la créativité lexicale (Guilbert, 1975).

La motivation est donc liée à la **compétence** et à la **conscience linguistiques** du locuteur. Elle sous-entend de la part de ce dernier aussi bien l'**étymologie** que la **grammaire** et plus précisément la **morphologie lexicale**. La motivation permet ainsi au locuteur natif de produire et naturellement de reconnaître des dérivés à partir d'un radical et au locuteur non natif de comprendre et, éventuellement, de produire une lexie nouvelle aussi complexe soit-elle, en partant des affixes dérivationnels, dès lors qu'il en connaît la valeur des éléments constitutifs. Entendue ainsi, cette motivation peut-être dite **morpholexicale**.

Cependant, il est important de souligner qu'en synchronie, des mots morpholexicale^{ment} dérivés peuvent ne plus avoir de liens de motivation du tout (par exemple entre «avenue» et «venir», entre «chanter» et «déchanter»). **Seul un effet de sens particulier**, voulu par exemple par le terminologue, peut ressusciter ce lien de motivation.

C'est en perdant ce lien de motivation que naît la démotivation d'une lexie; c'est en redonnant ce lien que nous aboutissons à la remotivation.

LA MOTIVATION CULTURELLE

Mais aussi «productive» qu'elle puisse paraître du point de vue structurel, la motivation morpholexicale ne peut être privilégiée dans notre démarche si elle n'intègre pas la saisie de la réalité socioculturelle dont la prise en compte permettra une meilleure adoption par la société impliquée.

Une autre motivation, non structurelle elle, peut donc être associée à la motivation morpholexicale. Elle serait définie comme l'ensemble des mémoires individuelles et de la mémoire sociale contenue dans un environnement dans lequel elle est entretenue. Seule cette motivation détermine le choix des symboles entrant dans la motivation. En somme, il s'agit d'une **motivation culturelle**, ce qui constitue une deuxième dimension de la mémoire des mots, puisqu'il se pose le problème de la **manière** dont cette mémoire est entretenue. En effet, la mémoire des mots fait appel à la fois au **contenu** (sens) et aux

contenants (mots, référents) **connus des locuteurs**. Il y a donc un autre problème qui se pose au terminologue, celui d'**espace** et de **support** (objets). Car il y a souvent un risque d'oubli des mots si les objets disparaissent.

Pour comprendre cette problématique rappelons l'origine du terme *ordinateur* racontée par Marcellesi (1979 : 176-177) : «en France, l'apparition des ordinateurs de première génération posa des problèmes de terminologie. Trois solutions étaient possibles : ou conserver l'un des deux termes anglo-américains; ou employer le terme *calculateur*, traduction de *computer* et employé jusque-là en mécanographie; ou créer un néologisme. C'est IBM FRANCE qui, en 1954, se saisit de ce problème. Elle ne voulait pas employer *calculateur* qui n'exprimait pas bien la révolution technique et la polyvalence des emplois de la machine. L'expression "*Machine électroniques à données processionnelles*"⁷ fut envisagée un moment, mais rapidement abandonnée. IBM FRANCE s'adressa alors à M. Jacques PERRET, professeur à la Faculté de Paris, qui proposa le terme ORDINATEUR, repris de l'expression théologique, "*Dieu, Grand ordinateur du monde*", expression dans laquelle "ordinateur" désigne Dieu comme celui qui met de l'ordre dans le monde selon un plan».

Ce qui est important de souligner ici, c'est que «ce terme désignait au XV^e siècle le prélat qui était chargé de calculer, selon un calendrier lunaire, la date annuelle de la fête de Pâques par rapport à laquelle toutes les célébrations du cycle liturgique devaient être ordonnancées. Cette fonction n'étant plus du tout pratiquée, le mot "ordinateur" [est] tombé en désuétude» (Diki-Kidiri et al. 1997 : 96).

De même «lorsqu'on a voulu traduire en français *software* et *hardware*, une quinzaine de paires de termes ont été proposées à l'Office de la langue française (OLF) dont *mentaille / quincaille*. Aucune d'elles n'avait été retenue. Mais leur étude a permis d'explicitier et d'affiner les critères d'acceptabilité dont la prise en compte a permis de trouver le couple *logiciel / matériel* qui sera consacré depuis lors» (Diki-Kidiri, 1997: 96-97).

Bien que certaines des propositions relevaient bien de la motivation morpholexicale interne à la langue française, par dérivation ou par composition, elles ont été écartées au profit d'une unité lexicale qui retourne aux sources culturelles, donc au passé.

Notre motivation sera ainsi plus socioculturelle contenant la dimension historique.

Nous définirons alors la **démotivation**, comme une **neutralisation du déjà connu**; tandis que la **remotivation** sera entendue comme **une réactivation du déjà oublié**; toutes deux manifestent de l'évolution de la langue et permettent de verbaliser les nouveaux objets (matériels ou épistémologiques).

L'une et l'autre recourent à ce que Pepermans (1991 : 21) appelle la «charge sémantique» qu'il définit comme «la possibilité virtuelle pour une unité lexicale d'acquérir un quantum de sémantité», «c'est-à-dire plus de sens» et la «charge notionnelle» entendue comme «la possibilité virtuelle pour une unité lexicale d'acquérir un quantum de

⁷ Rappelons que cette expression n'est qu'une traduction de l'américain : *Electronic Data Processing Machine* avant qu'il n'opte pour *Computer*.

notionalité», «c'est-à-dire tendre vers une seule notion». La «charge sémantique» et le «quantum de notionalité» faisant passer le terme au mot ou le mot au terme, offrent ainsi au terminologue la possibilité «d'étudier, d'un point de vue étymologique ou diachronique, le passage d'une unité lexicale de la langue courante à la langue spécialisée ou, inversement, de la langue spécialisée à la langue courante» Pepermans (1991 : 22). Et en l'occurrence de passer de la langue vieillie (*morte*) à la langue vivante.

QUELQUES EXEMPLES DE DÉNOMINATION CULTURELLE

Ainsi, quand il a fallu trouver un mot pour *dictionnaire* en bangala (langue véhiculaire du nord est du Congo), nous aurions pu «emprunter» la construction du lingala, **maloba ma lokóta**, littéralement les «mots de la langue» proposé par Everbroek⁸, ce qui n'est rien de moins qu'une adaptation des mots *vocabulaire* ou *lexique*. Aidé d'un des locuteurs le plus âgé dans son récit de vie, nous lui avons préféré le terme **ɓagó**.

En effet, selon cet informateur, **ɓagó** remplissait trois fonctions :

1° de par son âge ce *vieillard* était un *sage*, au jugement *impartial*; il pouvait donc régler les *litiges* entre les joueurs en cas de dispute;

2° on lui confiait aussi la *garde des enfants* pendant que les personnes valides étaient dans les travaux des champs;

3° c'est chez lui qu'on construisait le grenier⁹ dans lequel on *gardait les semences* (symbole de survie et de conservation) pour la saison agricole suivante;

4° par la suite **ɓagó**, perdant son trait +humain, a fini par désigner, par métaphore, le *grenier* dans lequel on gardait les semences.

N'étant plus d'usage, parce que remplacé par l'emprunt au français «conseiller», dans un des ses sens, nous l'avons ressuscité par «remotivation» pour dénommer l'objet dictionnaire.

Qu'est-ce, au fait, qu'un dictionnaire ? Dans Edema (1994 : 14), pour justifier ce choix de **ɓagó**, nous nous en expliquons ainsi :

«Instrument éminemment didactique, le dictionnaire a la prétention d'être l'outil le plus complet d'une langue puisque, au delà de son aspect purement linguistique, c'est toute la personnalité — culturelle, sociale, politique, scientifique, technique — d'un peuple qu'il reflète à travers elle. Il nous a paru que les sens différents du mot bangála, **ɓagó**, (**dépositaire, conseiller, grenier**) conviendrait à la définition que nous entendons du dictionnaire, dans la mesure où il nous semble parfaitement remplir ce triple rôle : oui le dictionnaire est bien le *dépositaire* du génie de la langue, parce qu'il est censé contenir toutes les unités lexicales de celle-ci; en tant que *conseiller* il permet de réconcilier les locuteurs en conflit avec la norme de la langue si tant est que cette norme existe; en tant que *grenier*, "étage supérieure" de la description linguistique, il conserve la langue dans tous ses mécanismes à l'abri des assauts du temps, c'est-à-dire à l'usure de la mémoire».

⁸ *Dictionnaire lingála-français, français-lingála; maloba ma lokóta lingála*, Kinshasa-Limete, Éditions de l'Épiphanie, 1985, 358 p.

⁹ Et de fait, le grenier ne conserve-t-il pas notre histoire ?

Un savant africain n'a-t-il pas dit que chaque fois qu'un vieillard meurt c'est une bibliothèque qui brûle ?

C'est en recourant aux lexies offrant une telle densité qu'on peut mieux dénommer, nous semble-t-il, car la démarche ne fait que reconstruire le schéma dont dispose déjà la langue, face au nouveau, pour appréhender l'inconnu, en s'appuyant sur le connu.

Revenons encore aux différentes dénominations du vélo présentées *supra*.

Chez les Kuba, seul *l'effet* a d'abord compté. Peuple d'artistes, ayant déjà une *longue expérience* des fresques, leur intérêt le plus immédiat s'est focalisé beaucoup plus sur *l'empreinte* de l'objet que sur son *identité*. Leur *passé artistique* qui constitue leur quotidien a pesé plus que l'utilité de l'objet.

Chez les Liko, c'est *l'attribut* que confère ce moyen de locomotion qui a présidé à la nomination de l'objet. En acquérant *deux jambes supplémentaires*, l'homme (re)devient *quadrupède*¹⁰ et par métonymie (transfert ou translation) c'est l'objet même qui se nomme «quatre-pieds». Il est inutile de souligner que ni l'homme, ni le vélo, séparé l'un de l'autre, n'ont quatre jambes.

Étymologiquement, qu'est-ce, en fait, qu'un «vélocipède», sinon «des pieds-rapides» ou «deux-pieds-qui-vont-vite» ? Et plus on a des «jambes», plus on va vite.

Chez les Centrafricains, c'est la *matière* d'une des composantes principales de l'objet qui a retenu le plus l'attention, suite à une des *conséquences les plus marquantes de la colonisation*. Ce n'est pas la réunion des «deux roues en caoutchouc» qui en font un vélo. Mais sans le caoutchouc le vélo ne saurait aller *véloce*ment. Il n'est donc pas un élément accessoire.

Au Mali le vélo est vu comme une *autre série* de cheval ou plus exactement comme une *variation d'un moyen de transport apparenté au cheval*. En effet, le «cheval de fer» n'est pas un cheval mais une spécification d'un moyen de transport ressemblant, à grands traits, au cheval. Seuls les traits qui établissent cette relation d'assimilation de ce qu'ils ont en commun les font coïncider. Ces traits sont les suivants :

- moyen de transport, comme le cheval;
- qui va aussi vite que le cheval;
- sur lequel l'homme se pose en califourchon, comme sur le cheval, en plus sur la même selle;
- en posant ses pieds sur des pédales, comme sur les étriers de l'harnais du cheval;
- le vélo a un guidon (motivation morpholexicale à partir du verbe «guider») qu'on manie comme les rênes du harnais¹¹.

¹⁰ Non pas qu'il retourne à l'enfance en marchant à quatre pattes, comme le veut la devinette.

¹¹ Il n'est pas inutile de rappeler que *rêne* était «en concurrence avec *guide* [qui a donné *guidon*] pour les chevaux attelés» (Cf. *Le Robert Dictionnaire historique de la langue française*).

Et ce n'est pas la différence de nombre de pieds (deux pour le vélo et quatre pour le cheval) qui empêchera d'assimiler l'un à l'autre, l'essentiel étant que la majorité de traits «s'enchâssent» du point de vue «reconceptuel».

À présent, mettons-nous dans les dispositions d'un locuteur lilikophone qui voit pour la première fois de sa vie un train. Après s'être fait expliquer son fonctionnement, il s'inquiète de savoir comment *ce long tas de ferraille* peut se retourner ou faire marche arrière, comme une voiture ou un camion le font facilement. Le «connaisseur» lui apprend qu'il n'a pas besoin de la même manoeuvre : il a deux «têtes», une à chaque bout; il suffit que le conducteur se mette à l'une d'elles pour le faire marcher dans un sens ou dans un autre. Faisant appel à ses connaissances, le locuteur lilikophone s'exclame : mais c'est un **mambélé-ma-suka** !

Mambélé-ma-suka, en langue liko, désigne tout simplement l'**amphisbène**, curieux reptile que le folklore de tous les pays a toujours défini comme un serpent «à deux têtes qui peut aller dans les deux directions». Non ce n'est pas un serpent et il n'a pas deux têtes. En fait on ne distingue pas, chez l'amphisbène, la queue de la tête. Et c'est vrai que ce reptile se déplace «tantôt en avant, tantôt en arrière, avec la même facilité» (Jared et al. 1977 : 38). «La taxonomie moderne les range dans un sous-ordre, séparé des squamates ou reptiles pourvus d'écailles, dénommé *Amphisbaenia*» (Jared et al. 1977 : 38).

L'amphisbène est dit **ikirumirhabiri** en kinyarwanda, «celui-qui-mord-de-deux-côtés¹²», parce qu'il a «deux têtes».

Quels sont les traits «objectifs» qui unissent le train et cette espèce de serpent dans ce cas-ci ? :

- ils sont tous les deux longiformes;
- quand il roule, le train paraît de loin souple et donc malléable, comme le serpent;
- ils empruntent facilement les courbes;
- le train glisse (rampe) comme le reptile qu'est le serpent;
- si le train a (souvent) une locomotive à chaque bout, le **mambélé-ma-suka** a une tête à chaque bout; c'est pourquoi on ne distingue pas la queue de la tête.

Si le terme du kinyarwanda est analysable, **i-ki-rum-ir-ha-biri**, *grosso modo* *qui-mord-deux*, on ne peut émettre que des hypothèses sur le terme du liliko, en recourant aux langues de la même famille comme le lingala et le kiswahili, ou à ce que les bantuistes appellent «bantu commun» (ensemble du fonds lexical reconstruit). On pourrait alors expliquer que dans «mambele» il y a *mbili* du kiswahili qui signifie *deux*. À moins que ce ne soit du proto-bantu ***bele** signifiant *avant, devant* et que «suká» s'apparente à **súka** du lingala signifiant *bout, fin*. Le composé signifierait «celui qui a deux bouts identiques» ou «celui qui a une tête par devant et une tête par derrière».

¹² Par métaphore, **ikirumirhabiri** désigne aussi un homme politique qui «mange à tous les râteliers». Au Congo-Kinshasa ce genre d'homme «perfide» est appelé «mouche» parce qu'avec ses six pattes, tout en se maintenant sur deux pattes du milieu, elle peut applaudir des pattes de devant tout comme de celles de derrière, sans que ceux qui la voient d'un côté ne se doutent de ce qu'elle fait de l'autre.

Quant à l'image du serpent que contient l'amphisbène, on peut se demander comment un symbole aussi répulsif pourrait être accepté par la société. «Heureusement», si l'on ose dire, il existe des «précédents» lexicaux impliquant le serpent et les autres moyens de transport au Congo Démocratique. Ainsi un avion de marque DC (Douglass Company) de type DC 8 était-il surnommé au Congo *serpent de l'air* puisque par son long fuselage il rappelle le serpent. De même un long camion de marque *Nissan* a-t-il reçu la dénomination de *serpent de route*. Dans un autre registre, un général de l'armée avait aussi reçu un sobriquet de *serpent de rail* tout simplement parce que, pour éviter les mines antichars posées par les rebelles sur les routes, ce rusé général les avait fait emprunter la voir ferrée... Le *serpent* comme train ne viendrait que compléter la série.

En conclusion de ces quelques exemples, nous terminerons par ce que la philosophie scolastique disait de la dénomination. Selon elle, la dénomination ne peut se faire que «par l'intermédiaire de la détermination d'une des propriétés» de l'objet en question. Aussi faut-il distinguer la **dénomination intrinsèque**, celle qui se conçoit de l'**intérieur**, dans la substance même, de la **dénomination extrinsèque**, celle qui prend sa source de l'**extérieur**, dans quelque chose qui n'est pas dans la substance de l'objet. C'est peut-être dans la dénomination extrinsèque que l'arbitraire du signe se conçoit mieux, parce qu'ici la «nomination» d'un objet (première dénomination, en tant que début du vocabulaire) n'est pas infuse ni mêlée à la substance (aucune association entre les mot *chaise* (français), *kiti* (en swahili), *fulikí* (liliko) et l'objet «chaise»). Elle est véritablement «arbitraire», ce que la dénomination, vue du point de vue culturel, n'est pas, ou du moins, ne devrait pas être. Et de fait, **dénommer** n'est-ce pas **intensifier** et **insister sur la nomination d'un objet** ?

En comparant les différents exemples en langues africaines on pourrait proposer comme première définition de la dénomination, vue par notre approche, en parodiant une définition célèbre de Jean Rostand de la «culture», comme étant «ce qui reste après avoir tout comparé». Ce sont les comparaisons qui sont les plus proches de la culture d'un peuple qu'il faudrait privilégier avant de recourir à la dérivation, à la composition, et en dernier lieu seulement, à la «solution la plus paresseuse» qu'est l'emprunt comme dit A. Rey.

Cependant, l'emprunt peut se remotiver ou se démotiver par rapport au sens d'origine. Dans ces conditions, il peut s'avérer fort instructif pour notre démarche. Il ne s'agit donc pas de rejeter tous les emprunts en bloc. En faisant un tri dans les mots empruntés, on se rendra compte qu'il y en a qui ont changé de référent.

En lingala par exemple, la lexie **zábulu**, issu du latin ecclésiastique *diabulus*, désigne certes le *diable* et a pour synonyme un autre mot emprunté au latin toujours **sátana** pour le même référent, mais la seconde lexie a tendance à supplanter la première pour désigner le «diable». Pourquoi ? Tout simplement parce que la première lexie signifie aussi *pellicule, négatif de photo*, sens que la seconde n'a pas.

Mais l'emprunt le plus proche sera toujours préférable. Ainsi en liliko, de trois signifiants concurrents désignant le *vélocipède* : **veló** (français), **kinga** (swahili) et **magu-mákwanganya**, le premier devrait-il être écarté au profit du deuxième si le troisième n'existait pas.

CONCLUSION

Quelle est la réaction de l'homme devant un objet nouveau ? Telle était la question à laquelle nous avons tenté de répondre à travers ces quelques lignes.

Cet exposé visait à explorer la voie la plus proche par laquelle passe le locuteur, quand il veut ignorer la traduction et l'emprunt, ce dernier n'étant qu'un pis aller en terminologie, mais aussi les possibilités qui s'offrent au linguiste s'il veut dénommer les nouvelles notions ou les nouvelles fonctions dans une langue quelconque. En effet, même si les langues se débrouillent toutes seules, le terminologue doit lui donner un coup de pouce. Cet exercice ne se limite donc plus au niveau des mots (par métaphores, dérivation), mais recourt aussi aux symboles des objets culturels. Ce qui rend la dénomination moins arbitraire que tout autre signe linguistique. Du coup elle ne se réalise qu'après l'opération déconstruction-reconstruction de l'objet nouvellement entré dans l'univers (ou dans sa culture) du locuteur. C'est cet acte que M. Diki-Kidiri (ici même) appelle «reconceptualisation».

En puisant dans ce que la langue et l'expérience ont déjà montré, nous voulons dire par là ce que la culture et les réserves morpholexicales d'une langue ont déjà donné comme illustrations, le terminologue paraît ne rien inventer, sinon emprunter un chemin déjà tracé. Dès lors la mémoire des mots dont nous traitons ici résiderait-elle dans le sens étymologique du mot composé *déjà* qui indique «un moment du présent ou du passé» ou «qu'une information appelée était connue mais oubliée» (Cf. *Le Robert, Dictionnaire historique de la langue française*) ?

La démotivation et la remotivation ne seraient-elles alors qu'un exercice de réminiscence ? C'est bien possible, car s'il faut en croire Barthes (1970 : 198), en matière de créativité lexicale «tout existe déjà, il faut seulement le retrouver : c'est une notion plus "extractive" que "créative"».

Comment définir alors la «mémoire des mots» ? Vogue-t-elle entre synchronie et diachronie ? La question est de savoir si la dénomination ne fait qu'extraire soit dans le réservoir de ressources sémantiques des mots existants, soit dans les structures linguistiques ou soit dans le culturel ou dans les trois à la fois.

La réponse peut être proposée comme suit : c'est ce qui préside à la sélection des traits aussi bien linguistiques que socioculturels qui détermine la dénomination. Oui la mémoire des mots réside bien dans le **déjà**, dans le **passé** et le **présent** : **passé de la culture et présent de la connaissance**, entre étymologie et créativité lexicale. La métallurgie traditionnelle du fer par exemple, bien que sa technique ait été oubliée, a laissé, chez les Liko, le mot **ábumá**, *scorie, gangue*, pour désigner les pièces de monnaie, là où le kiswahili a recouru à une motivation morpholexicale, **kichele** (du verbe *kuchelewa rester invendu, ne pas trouver preneur*; «*coiffer la Sainte Catherine*» pour une jeune fille); de même la table d'opération est-elle nommée **mundagi** chez les Liko toujours, lequel désignait le tronc de bananier sur lequel opérait le «cironciseur» traditionnel...

En recourant aux objets non courants du passé, nous avons donc affaire à une double mémoire : celle des mots et celle des techniques, celle des structures linguistiques

et celle de l'histoire, car, ne l'oublions pas, même «les mots les plus frustes vivent secrètement : leur passé vibre en eux. Qu'on les éveille et les dérange, on s'aperçoit que le passé de la langue est plus inattendu que son présent, trop souvent tissé de répétition et d'amnésie» (Rey, 1996 : 12). La mémoire des mots est alors une paléontologie.

RÉFÉRENCES

- ADDA, R. et al. (1979) : *Néologie et lexicologie. Hommage à Louis Guilbert*, Paris, Larousse, 224 p.
- ARRIVE, M., GADET, F. et M. GALMICHE (1986) : *La grammaire d'aujourd'hui. Guide alphabétique de linguistique française*, Paris, Flammarion, 720 p.
- BARTHES, R. (1970) : «L'ancienne rhétorique : aide-mémoire», *Communications* 16, Paris, Écoles des Hautes Études en Sciences Sociales, Seuil, pp. 172-223.
- CAPRILE, J.P. (dir) (1979) : *Contacts de langues et contacts de cultures 3. La créativité lexicale spontanée en Afrique Centrale par emprunt au français*. Paris, LACITO-Documents Afrique 3, 138 p.
- CELTA (1974) : *Actes du 1^{er} Séminaire des linguistes du Zaïre*, Lubumbashi, 501 p.
- CLAS, A. (dir) (1985) : *Projets lexiques spécialisés (LEXIS) et dictionnaires monolingues (DIMO). Guide de recherche en lexicographie et en terminologie*, Paris, ACCT, 158 p.
- DIKI-KIDIRI, M. (1996) : «La métaphorisation comme base culturelle de conceptualisation et source de néologismes terminologiques» Kacouri et al., *Questions de glottopolitique : France, Afrique, Monde méditerranéen*, Rouen, URA CNRS 1164, Université de Rouen, pp. 187-194.
- DIKI-DIKIRI, M., MBOJ, C. et A.B. EDEMA (1997) : «Des lexiques en langues africaines (sängö, wolof, lingála) pour l'utilisateur de l'ordinateur», *META, numéro spécial. Lexicologie et terminologie*, vol. 42, n° 1, pp. 94-109.
- DIOP, Cheik Anta (1975) : «Comment enraciner la science en Afrique : exemple wolof (Sénégal)», *Bulletin de l'Institut fondamental d'Afrique noire*, série B, Tome 37, n° 1, pp. 154-233.
- EDEMA, A.B. et P. EPANGA : «La diffusion du manioc en Afrique centrale», texte inédit, présenté lors de la «Table ronde» sur «Plantes, paysages et histoire en Afrique subsaharienne», Paris, mai 1994.
- EDEMA, A.B. et K. NDUKU (1993) : «Les variantes graphiques de l'IFA : un point de vue zaïrois», *Inventaires des usages de la francophonie. Nomenclature et méthodologies*, Paris, AUPELF-UREF/John Libbey, Actualité Scientifique, pp. 355-372.
- EDEMA, A. B. (1994) : *Dictionnaire bangála - français - lingála*, Paris, ACCT-SEPIA, 239 p.
- Encyclopédie philosophique universelle* (1990) : Paris, PUF, Tome 1, Les notions philosophiques.

- FOUCAULT, M. (1990) : *Les mots et les choses. Une archéologie des sciences humaines*, Paris, France Loisirs, 574 p.
- GUILBERT, L. (1975) : *La créativité lexicale*, Paris, Larousse, coll. «Langue et Langage», 285p.
- JARED, C., BURMAN, A. et N. LEON (1997) : «Les amphibènes, folklore et biologie», *La Recherche* 302, Paris, Société d'éditions scientifiques, pp. 38-40.
- KAMBA, M. (1981) : *Les formes verbales négatives dans les langues bantoues*, Tervuren, Musée Royal de l'Afrique Centrale, 350 p. + 23 p. hors texte.
- KLEIBER, G. (1984) : «Dénomination et relations dénominatives», *Langages* 76 : *La dénomination*, Paris, Larousse, pp. 77-94.
- MARCELLESI, C. (1979) : «Retour aux sources : quelques aspects du vocabulaire de l'informatique», Adda et al., *Néologie et lexicologie*, Paris, Librairie Larousse, coll. «Langue et langage», pp. 176-183.
- MORTON, J. (1984) : «La dénomination», *Langages* 76 : *La dénomination*, Paris, Larousse, pp. 19-30.
- MORTUREUX, M.-F. (1984) : «La dénomination, approche socio-linguistique», *Langages* 76 : *La dénomination*, Paris, Larousse, pp. 95-112.
- NYEMBWE, N. (1992) : «Terminologie et développement linguistique au Zaïre», *Terminologies nouvelles* 3, Québec, ACCT, Communauté française de Belgique, pp. 80-83.
- PALUKU, M. (1989) : «Méthodes de travail en terminologie au Zaïre», *Terminologies nouvelles* 3, Québec, ACCT, Communauté française de Belgique, pp. 37-41.
- PEPERMANS, R. (1991), «L'axe sens-notion : schéma d'interprétation théorique portant sur les rapports entre langue courante et langue spécialisée», *Terminologies nouvelles* 6, Paris-Bruxelles, ACCT-Communauté française de Belgique, Marc Van Campenhout éditeur scientifique, pp. 21-23.
- REY, A. (1979) : *La terminologie, noms et notions*, Paris, PUF, coll. «Que sais-je ?», 128 p.
- REY, A. (1996) : *Le réveille-mots. Une saison d'élection*, Paris, Éditions du Seuil, coll. «Point Virgule», 238 p.

Achévé d'imprimer par la SOTEPA

Dépôt légal : 3ème trimestre 1998

Réalisation

S E R V I C E D

12, Av. de la Liberté - El Menzah V - 2080 Ariana - Tunis Tunisie
Tél.: 232.649 / 752.889 - Fax : 751.914

La collection **Universités francophones**, créée en 1988 à l'initiative de l'AUFPELF-UREF, propose des ouvrages de référence, des manuels spécialisés et des actes de colloques scientifiques aux étudiants de 2ème et 3ème cycles universitaires ainsi qu'aux chercheurs francophones et se compose de titres originaux paraissant régulièrement.

Leurs auteurs appartiennent conjointement aux pays du Sud et du Nord et rendent compte des résultats des recherches et des études récentes entreprises en français à travers le monde. Ils permettent à cette collection pluridisciplinaire de couvrir progressivement l'ensemble des enseignements universitaires en français.

Enfin, la vente des ouvrages à un prix préférentiel destiné aux pays du Sud tient compte des exigences économiques nationales et assure une diffusion adaptée aux pays francophones.

Ainsi, la collection **Universités francophones** constitue une bibliothèque de référence comprenant des ouvrages universitaires répondant aux besoins des étudiants de langue française.

Cet ouvrage représente les Actes des Vèmes Journées Scientifiques du Réseau Thématique «Lexicologie, Terminologie, Traduction» de l'AUFPELF - UREF, organisées, en collaboration avec l'Association Tunisienne de Linguistique, à l'Université de Tunis I du 25 au 27 septembre 1997, autour du thème *La Mémoire des Mots*.

Les contributions que renferme cet ouvrage sont la preuve incontestable de la naissance d'un concept linguistique nouveau «La Mémoire des Mots» qui, de l'avis de tous les participants, a permis de cerner avec beaucoup plus de précision des aspects linguistiques jusque-là ignorés, et de créer une ouverture épistémologique favorisant le décloisonnement entre les disciplines et les spécialités pointues. Ainsi dans l'étude du même phénomène linguistique, le figement lexical par exemple, les éclairages de l'analyse morphologique, syntaxique ou sémantique se trouvent-ils complétés et enrichis par ceux du traitement informatique.

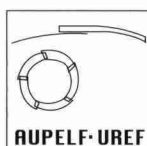
Multidisciplinarité, diversité des approches, réflexion théorique, traitement automatique du langage et culture sont autant de facettes traitées dans cet ouvrage qui apporte plusieurs réponses à des problèmes théoriques et des besoins de nature pratique en matière de recherche linguistique (traduction, terminologie, etc.).

Spécificités linguistiques et innovations technologiques, grâce à *La Mémoire des Mots*, participent ainsi à l'ouverture du même horizon épistémologique.

140,000 FF

40,00 FF - Prix préférentiel : Afrique, Asie, Amérique du Sud, Haïti, Maghreb

5,000 D.T. - Prix préférentiel étudiants



S E R V I C E D

TUNIS - TUNISIE