

Actes des
8^e Journées
scientifiques
du Réseau
de chercheurs
*Lexicologie,
terminologie,
traduction*

Lisbonne,
15-17 octobre 2009

Sous la direction de:

Marc
Van Campenhoudt
Teresa Lino
Rute Costa

PASSEURS DE MOTS, PASSEURS D'ESPOIR

Lexicologie,
terminologie
et traduction
face au défi
de la diversité



Passeurs de mots, passeurs d'espoir **lexicologie, terminologie et traduction** **face au défi de la diversité**

Actes des Huitièmes Journées scientifiques
du Réseau de chercheurs
Lexicologie, terminologie, traduction

Lisbonne, 15-17 octobre 2009

Sous la direction de :

Marc Van Campenhout

Institut supérieur de traducteurs et interprètes, Haute École de Bruxelles

Teresa Lino

Rute Costa

Universidade nova de Lisboa

Textes préédités par Nathalie Lemaire

COMITÉ D'HONNEUR

Michèle GENDREAU-MASSALOUX
Recteur honoraire de l'Agence universitaire de la Francophonie,
Responsable Formation, Enseignement supérieur, Recherche
de la Mission Union pour la Méditerranée
Bernard CERQUIGLINI
Recteur de l'Agence universitaire de la Francophonie
Bernard QUEMADA
Directeur d'Études honoraire à l'EPHE, Directeur de recherche émérite au CNRS
Fernando RAMÔA RIBEIRO
Président Fondateur de l'*Associação Portuguesa dos Doutorados em França* - APDF

COMITÉ SCIENTIFIQUE

Ibrahim BEN MRAD
Université de Tunis, La Manouba
Xavier BLANCO
Université autonome de Barcelone
Mame Thierno CISSÉ
Université Cheikh Anta Diop, Dakar
Manuel Célio CONCEIÇÃO
Université d'Algarve
Rute COSTA
Université nouvelle de Lisbonne
Isabel DESMET
Université Paris 8
Marie-Claude L'HOMME
Université de Montréal
Teresa LINO
Université nouvelle de Lisbonne
Franck NEVEU
Université de Caen Basse-Normandie
Philippe THOIRON
Université de Lyon 2
Marc VAN CAMPENHOUDT
Institut supérieur de traducteurs et interprètes, Haute École de Bruxelles

COMITÉ D'ORGANISATION

Manuel Célio CONCEIÇÃO
Université d'Algarve
Rute COSTA
Université nouvelle de Lisbonne
Teresa LINO
Université nouvelle de Lisbonne
Isabelle OLIVEIRA
Université Paris 3
Raquel SILVA
Université nouvelle de Lisbonne

Copyright © 2011 Éditions des archives contemporaines et en partenariat avec l'Agence universitaire de la Francophonie

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

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

ISBN : 9782813000521

Les textes publiés dans ces actes ont été sélectionnés par le comité scientifique du colloque. Leur contenu n'engage toutefois que la responsabilité de leurs auteurs. Pour faciliter la lecture, la mise en page a été harmonisée, mais la spécificité de chacun, dans le système des titres, le choix de transcriptions et des abréviations, l'emploi de majuscules, la présentation des références bibliographiques, etc. a été le plus souvent conservée.

Nous tenons à remercier Nathalie Lemaire pour son travail de préédition et Serge Paulus pour la mise en page de ce volume.

Index des auteurs

| | |
|---|-----|
| <i>Alic, Liliana</i> | 299 |
| <i>Alves, Ieda Maria</i> | 367 |
| <i>Anastassiadis-Syméonidis, Anna</i> | 287 |
| <i>Anckaert, Philippe</i> | 549 |
| <i>Arnaud, Pierre J.L.</i> | 503 |
| <i>Bacquelaine, Françoise</i> | 235 |
| <i>Baptista, Jorge</i> | 97 |
| <i>Berkai, Abdelaziz</i> | 537 |
| <i>Bevilacqua, Cleci Regina</i> | 219 |
| <i>Bordet, Geneviève</i> | 579 |
| <i>Bouchaddakh, Samia</i> | 63 |
| <i>Bramati, Alberto</i> | 479 |
| <i>Calberg-Challot, Marie</i> | 567 |
| <i>Cartier, Emmanuel</i> | 619 |
| <i>Català, Dolors</i> | 97 |
| <i>Chetouani, Lamria</i> | 311 |
| <i>Chouma, Valentina</i> | 287 |
| <i>Conceição, Manuel Célio</i> | 211 |
| <i>Contente, Madalena</i> | 557 |
| <i>Daille, Béatrice</i> | 605 |
| <i>Damas, Luc</i> | 567 |
| <i>Dancette, Jeanne</i> | 161 |
| <i>de Caldas, Sandra</i> | 341 |
| <i>de Vecchi, Dardo</i> | 149 |
| <i>Doll, Frédéric</i> | 441 |
| <i>Drouin, Patrick</i> | 441 |
| <i>Dziadkiewicz, Aleksandra</i> | 23 |
| <i>El Mountassir, Abdallah</i> | 277 |
| <i>Eyckmans, June</i> | 549 |
| <i>Feldekircher Müller, Alexandra</i> | 219 |
| <i>Fréjaville, Rosa Maria</i> | 593 |
| <i>Goeuriot, Lorraine</i> | 605 |

| | |
|--|----------|
| <i>Gołkowska, Kaja</i> | 139 |
| <i>Guidère, Mathieu</i> | 259 |
| <i>Humbley, John</i> | 325 |
| <i>Jacquet-Pfau, Christine</i> | 325 |
| <i>Kacprzak, Alicja</i> | 355 |
| <i>Krieger, Maria da Graça</i> | 219 |
| <i>Kübler, Natalie</i> | 579 |
| <i>L'Homme, Marie-Claude</i> | 651 |
| <i>Ledouble, Hélène</i> | 633 |
| <i>Lerat, Pierre</i> | 177 |
| <i>Mangeot, Mathieu</i> | 35 |
| <i>Maniez, François</i> | 671 |
| <i>Massoussi, Taoufik</i> | 185, 245 |
| <i>Mejri, Salah</i> | 197 |
| <i>Mejri, Soumaya</i> | 245 |
| <i>Mitsiaki, Maria</i> | 287 |
| <i>Mogorrón Huerta, Pedro</i> | 517 |
| <i>Nakao, Yukie</i> | 605 |
| <i>Navlea, Mirabela</i> | 375 |
| <i>Nguyen, Hong-Thai</i> | 35 |
| <i>Oliveira, Isabelle</i> | 429 |
| <i>Oliver González, Antoni</i> | 419 |
| <i>Pecman, Mojca</i> | 579 |
| <i>Petit, Gérard</i> | 185 |
| <i>Pimentel, Janine</i> | 651 |
| <i>Ploux, Sabine</i> | 429 |
| <i>Pognan, Patrice</i> | 83, 399 |
| <i>Raharinirina Rabaovololona, Lucie</i> | 469 |
| <i>Roche, Christophe</i> | 567 |
| <i>Sablayrolles, Jean-François</i> | 325 |
| <i>Saïd, Mosbah</i> | 455 |
| <i>Segers, Winibert</i> | 549 |
| <i>Serhan, Carla</i> | 111 |
| <i>Silva Filho, Sebastião</i> | 557 |

| | |
|-------------------------------------|-----|
| <i>Simunic, Zrinka</i> | 123 |
| <i>Täïf, Miloud</i> | 83 |
| <i>Todiraşcu, Amalia</i> | 375 |
| <i>Trurnit Verbic, Renate</i> | 633 |
| <i>Valli, André</i> | 479 |
| <i>Vàzquez Garcia, Mercè</i> | 419 |
| <i>Verlinde, Serge</i> | 389 |
| <i>Verplaetse, Heidi</i> | 549 |
| <i>Vletsi, Eleni</i> | 287 |
| <i>Vu, Van Dai</i> | 493 |
| <i>Wermuth, Cornelia</i> | 549 |
| <i>Zotti, Valeria</i> | 49 |

Séance d'ouverture

Allocution d'ouverture

Marc Cheymol,
Directeur,
Langue française, diversité culturelle et linguistique,
Agence universitaire de la Francophonie.

Monsieur le Vice-Recteur de l'Université nouvelle de Lisbonne,

Monsieur le Sous-Directeur pour la recherche de la Faculté des sciences sociales et humaines,

Madame la Conseillère de coopération et d'action culturelle de l'Ambassade de France,

Madame la Chef de la mission Développement et modernisation de la langue française, de la Délégation générale à la langue française et aux langues de France,

Monsieur le Président de l'Association portugaise des docteurs diplômés en France,

Monsieur le Coordonnateur du Réseau LTT,

Mesdames et Messieurs,

Chers Collègues, Étudiants, Enseignants et Chercheurs,

Je voudrais tout d'abord, au nom de l'Agence universitaire de la Francophonie, remercier publiquement les autorités universitaires de l'Université nouvelle de Lisbonne pour l'accueil chaleureux qui nous a été réservé et auquel nous sommes très sensibles. Je suis heureux de me trouver dans cette ville où j'ai déjà eu l'occasion de venir à titre personnel, bien sûr, car c'est une étape obligée dans la formation de tout Européen, mais aussi en tant que représentant de l'AUF : nous y comptons une université membre et y avons des perspectives de développement et de soutien de l'enseignement/apprentissage du français non seulement dans les universités de Lisbonne, mais aussi de tout le Portugal.

Je remercie le professeur Marc Van Campenhoudt, Coordonnateur du Réseau LTT, et avec lui tout le comité scientifique du Réseau LTT d'avoir eu l'initiative de ces journées et d'y avoir associé l'Agence universitaire de la Francophonie. J'adresse en particulier un sentiment de vive reconnaissance à la professeure Maria Teresa Lino, qui a su mobiliser autour de ce projet les forces vives de son université. Je profite de l'occasion pour remercier toute son équipe qui s'est dépensée sans compter pour la mise en œuvre d'une manifestation aussi importante.

Je représente ici le recteur de l'AUF, M. Bernard Cerquiglini, dont je vous transmets les salutations et les vœux pour la réussite de vos travaux, mais aussi les regrets, car il n'a pas pu se libérer de ses obligations pour vous accompagner comme il l'avait souhaité dans ces journées, et partager l'amitié et les compétences qui le lient à beaucoup d'entre vous.

Dire qu'une manifestation est importante, c'est une platitude, excusez-moi, c'est utiliser un terme bien galvaudé et si je me suis permis de le faire c'est pour pouvoir illustrer brièvement pourquoi. Elle est importante pour le Réseau LTT, elle est importante pour l'AUF et elle est importante pour la francophonie.

LTT

Il y a toujours eu un lien privilégié, un lien organique – pour ne pas dire un cordon ombilical – entre l'Agence universitaire de la francophonie et LTT. J'aimerais insister brièvement, afin de procéder à l'éclairage institutionnel, qui me revient, sur le rôle d'interfaces qu'ont joué les réseaux entre les universités et l'AUF en tant qu'opérateur universitaire de l'Organisation internationale de la Francophonie.

La manifestation qui s'ouvre aujourd'hui se situe à un moment crucial de l'histoire de notre institution, et partant, de votre réseau. En réalité nous sommes réunis ici pour une fête. Une fête de l'esprit et de la langue – les journées scientifiques du Réseau LTT sont toujours une fête, elles sont ressenties comme cela, je le sais –, mais aussi une commémoration. Vous l'avez remarqué, LTT fête ses vingt ans (ou vingt et un, car comme les millénaires, c'est une date sentie comme si importante que l'on se bat pour savoir quel point de départ on choisit), c'est le fruit d'une longue histoire depuis 1989. Il a été le premier réseau de l'AUF ; aujourd'hui, c'est le plus grand des réseaux du programme avec ses 670 membres qui représentent près de 40 % du nombre total des chercheurs rattachés au champ disciplinaire *Langue française, diversité culturelle et linguistique*.

Une fête, un anniversaire en particulier, c'est toujours une charnière, une articulation, un entre-deux. Lorsqu'on vient de souffler ses vingt bougies, on a encore un peu 19 ans et l'on n'a pas encore tout à fait 20 ans. C'est un moment infime où l'on bascule dans la décade suivante. C'est aussi le lieu, et le moment, d'un passage. On est tenté de regarder vers le passé, mais c'est dans l'avenir qu'il faut se jeter. Dans le cadre de la réorganisation de l'AUF dont j'ai parlé, LTT est à un tournant de son histoire.

Francophonie

La fête a été placée sous le signe de la diversité. Je relève que les 60 communications sélectionnées proviennent d'une vingtaine de pays : Algérie, Belgique, Brésil, Cameroun, Canada, Côte d'Ivoire, Croatie, Espagne, France, Grèce, Italie, Liban, Madagascar, Maroc, Pologne, Portugal, Roumanie, Sénégal, Suisse, Tunisie, Vietnam. Vous avez reconnu dans cette liste non seulement des pays de la francophonie,

et des pays de langue portugaise, d'Europe et d'Amérique, mais aussi d'autres pays européens, en particulier de la Méditerranée, qui prouvent ici leur francophonie.

Ainsi, vos journées scientifiques célèbrent aussi la diversité linguistique. Elles rassemblent des chercheurs issus d'une grande variété de pays francophones, voire de territoires où le français est une langue seconde, sans statut officiel ni lien génétique ou historique avec les langues locales. « Par-delà l'hommage rendu aux nombreux chercheurs lusophones actifs au sein du réseau, le choix de Lisbonne comme lieu des 8^{es} *Journées scientifiques* illustre la dynamique d'une francophonie nouvelle, vécue comme un espace de rencontre et de partage. »

À cette diversité géographique et linguistique, vraiment internationale, qui dépasse le cadre de la seule francophonie plurilingue, s'ajoute une diversité disciplinaire pour permettre à des chercheurs confirmés de « revisiter, dans leurs champs d'investigations respectifs les théories et paradigmes autour du multilinguisme, posant les questions de la transmission des connaissances et de l'intercompréhension entre toutes les langues en présence dans les aires linguistiques des chercheurs francophones, hispanophones, lusophones ou anglophones. » (Je reprends ici les termes de la problématique du colloque.)

« La question de la diversité est évoquée résolument par le mot *traduction*. Trop longtemps considérée comme un pis-aller, voire une capitulation, la traduction contribue à la circulation des idées. Elle permet d'appréhender l'autre, de connaître ses travaux et de s'inscrire dans un débat. Aujourd'hui, un état des lieux lucide conduit à préférer un développement harmonieux des langues à une compétition entre langues de grande expansion. La langue n'est plus un pré carré à préserver, elle est le lieu de la rencontre, du métissage, de l'émergence des savoirs. »

La diversité est bien un « défi ». En l'associant à un mot qui exprime le risque, vous avez dans votre titre redonné du sens à ce terme usé de « diversité », dévalué par une utilisation excessive dans les discours politiques et dans les médias. C'est pourtant le défi qui nous a été lancé en constituant le programme que je dirige et en m'en confiant la responsabilité. C'est le défi de la Francophonie.

Lorsque la Francophonie en 2005 – année de vos précédentes journées scientifiques à Bruxelles – définit ses missions dans son cadre stratégique décennal, l'AUF, comme les autres opérateurs de la Francophonie, se dote d'une programmation quadriennale répondant à ces objectifs. Par souci de simplification et de cohérence, elle réduit le nombre de ses actions et les regroupe en 5 programmes dont le premier est intitulé *Langue française, diversité culturelle et linguistique*. Par ailleurs, elle déconcentre ses activités vers les bureaux régionaux de manière à se rendre plus proche des besoins du terrain et plus attentive aux défis mondiaux.

Aujourd'hui, l'AUF ouvre une nouvelle période quadriennale avec la volonté de répondre aux leçons tirées de la programmation 2005-2009. Je dirai un mot de

cette réforme, car elle vous concerne tous en tant que partenaires ou utilisateurs des programmes de l'AUF.

AUF

En prenant les rênes de l'AUF, il y a déjà près de deux ans, le recteur Bernard Cerquiglini a souhaité une réorganisation qui, tout en allant dans le sens de l'action de sa prédécesseure Michèle Gendreau-Massaloux, donne à l'institution plus de visibilité internationale afin d'assurer une meilleure compétitivité dans un monde concurrentiel. Cette réorganisation à plusieurs niveaux et qui se fait en plusieurs étapes – elle a été spatiale et administrative, elle est encore administrative et scientifique, et communicationnelle – a été approuvée par les conseils de nos instances en avril et décembre 2008, en avril 2009 et surtout en mai 2009 par l'Assemblée générale. Elle s'appuie sur trois principes : la simplification et visibilité de nos outils et de nos procédures, la priorité des contenus scientifiques sur les divisions administratives, la conformité des actions avec le double mandat d'opérateur multilatéral et d'association de 710 établissements membres.

Dans ce cadre, tous les outils de l'AUF (les réseaux de chercheurs, les filières francophones, les bourses, les programmes de coopération scientifiques inter-universitaires) seront remis à plat afin d'en alléger encore les coûts de fonctionnement (alléger les coûts de fonctionnement est le moyen de maintenir les crédits alloués aux programmes) et surtout de rendre notre institution plus moderne et plus apte à capter de nouveaux financements, l'AUF s'étant lancée dans une importante campagne de levée de fonds.

Nous sommes arrivés à une étape décisive de cette réforme, avec l'élaboration de la programmation de l'AUF pour les 4 années à venir, qui compte tenu du calendrier décennal qui nous est imposé par la Francophonie institutionnelle, sera proposée aux instances de la Francophonie à la fin de l'année et qui devrait présenter un visage entièrement rénové de l'AUF.

L'AUF se propose, dans sa programmation 2010-2013, de renforcer son soutien à ses membres, en recentrant ses actions sur les missions et les expertises de l'AUF, et en poursuivant la déconcentration de ses activités vers le terrain. Elle mettra en place une veille sur le monde universitaire et renforcera également son engagement aux objectifs de la Francophonie. Pour cette nouvelle programmation, et à partir de ses missions propres, l'AUF a déterminé trois objectifs stratégiques à atteindre dans les quatre années à venir :

- Soutenir les stratégies de développement scientifique des établissements membres ;
- Rendre la communauté scientifique francophone indispensable sur la scène internationale ;

- Faire émerger une génération d'enseignants, de chercheurs, d'experts et de professionnels, acteurs du développement.

On voit bien comment une manifestation comme celle-ci répond au moins à deux de ces objectifs. L'AUF revoit ses modalités d'action en développant une approche par projet et en mettant en place un système de suivi de qualité et de mesure de l'impact. Chaque projet répondra à un ou plusieurs de ces objectifs. Il fera converger dans une véritable synergie les différents outils de l'AUF : actions de recherche, mobilités, horizons francophones, formations ouvertes et à distance, programmes de coopération interuniversitaire. Il privilégiera des projets d'appui à la formation, à la formation à la recherche, de préférence à des événements ponctuels.

Dans ce contexte, on parle d'une transformation des réseaux de chercheurs sous leur forme actuelle. On a souligné la force de certains d'entre eux, comme précisément LTT, leur crédibilité auprès des établissements d'enseignement supérieur et leur présence symboliquement forte dans la communauté scientifique. On a insisté sur le nécessaire ancrage régional des projets, qui doivent répondre aux besoins locaux. Certaines forces reposent sur les établissements, les disciplines, les associations, et les synergies doivent être maintenues et enrichies avec ces différents niveaux d'action. Les réseaux évolueront vers une nouvelle forme d'organisation de la communauté scientifique, capable de répondre plus clairement aux objectifs tracés par les instances de la Francophonie et aux besoins de développement des établissements membres de l'AUF.

Conclusion

La fête, la commémoration est donc bien un passage, voire « un rite de passage » – c'est le titre de l'une de vos communications. La conception des réseaux de chercheurs et de leurs manifestations scientifiques tels que vous les avez connus depuis vingt ans sera bientôt dépassée. LTT prouvera, par cette manifestation qui commence maintenant, mais je crois nécessaire que vous l'ayez tous présent à l'esprit dans vos travaux, qu'il est capable de passer à l'âge adulte et de répondre à la fois aux objectifs de la francophonie et aux objectifs des membres de l'AUF, même dans une nouvelle configuration de cette institution.

Lisbonne n'est-elle pas le lieu idéal de ce passage ? Le Portugal est aussi un symbole de l'articulation, lieu de croisements et de passage. Il a une situation stratégique unique : pays atlantique, tourné vers l'Amérique, l'Extrême-Occident, l'Afrique, les Indes orientales, il est aussi (même s'il est difficile pour moi de le dire) un pays méditerranéen par sa culture et sa participation à la culture de la Méditerranée. D'ailleurs, il fait politiquement partie de l'Union pour la Méditerranée. Il est aussi une articulation entre le Nord et le Sud : il est tourné vers le Sud, mais son appartenance à l'Europe en fait un pays du Nord, au nord du Sud et au sud du Nord. Il est aussi un lieu privilégié de l'interlinguistique, et un endroit où la francophonie mérite d'être soutenue et défendue.

J'ai parlé de notions « dépassées », au risque de faire sursauter certains, pour faire passer le message de la nouvelle AUF, et pour projeter LTT vers l'avenir, comme d'autres ont parlé de « déconstruire » lorsqu'ils entendaient par là « construire ». Il faut dépasser une certaine vision de la francophonie refermée sur son pré carré, comme il faut dépasser une conception devenue caduque de la coopération universitaire en francophonie. Il faut dépasser surtout les réactions de peur devant le changement et de repli sur soi. Je suis convaincu que LTT a beaucoup à y gagner. Le passage sera peut-être un passage difficile, mais il est nécessaire et présente de nouveaux avantages. Je voudrais aussi être moi aussi, sur l'avenir de LTT, un passeur non seulement de mots, mais d'espoir.

Je vous remercie.

Allocution d'ouverture

Marc Van Campenhoudt,
Coordonnateur du Réseau LTT

Monsieur le Directeur du Programme *Langue française, diversité culturelle et linguistique* de l'Agence universitaire de la Francophonie,

Monsieur le Vice-Recteur de l'Université nouvelle de Lisbonne,

Madame la Représentante de l'Ambassade de France,

Madame la Représentante de la Délégation générale à la langue française et aux langues de France,

Monsieur le Président de l'Association portugaise des docteurs diplômés en France,

Madame la Présidente du Comité organisateur,

Chères Collègues,

Chers Collègues,

Passeurs de mots, passeurs d'espoirs... Lorsque le comité scientifique s'est penché sur le choix d'une thématique pour ces *Huitièmes Journées scientifiques*, il était déjà exposé à la question de la diversité linguistique et culturelle : comment serait perçu ce « passeur » ? Traverse-t-il clandestinement les mers avec son lot de collègues du Sud, fait-il franchir des zones dangereuses à des terminologues en rupture de ban, vend-il sous le manteau d'illicites produits destinés à nos membres en manque d'inspiration ou fait-il, comme Charon, traverser les eaux du Styx aux âmes défunes du cercle des linguistes et des traducteurs disparus ?

Nul doute que nos travaux nous montreront que les passeurs sont ceux qui, comme le présentait notre appel à communications, concourent à répondre aux défis de la diversité linguistique. Les manières de répondre à ces défis sont multiples au sein de notre communauté, mais se font toujours écho sous les voûtes que soutiennent nos trois piliers : *lexicologie, terminologie, traduction*. Nous attendrons la synthèse finale pour aborder la délicate problématique de l'espoir qui naît de la diversité linguistique. Les réponses éventuelles se nourriront du contenu des nombreuses interventions auxquelles la thématique générale servira de toile de fond.

Nous aimons à dire que LTT est plus qu'un réseau : c'est une communauté de chercheurs, pour ne pas dire une famille élargie, au sein de laquelle nous aimons nous retrouver et accueillir et où nous sommes nombreux à avoir fait nos premières

armes. Les jeunes chercheurs sont encouragés à présenter des communications et associés aux actions de recherche, des formations sont systématiquement proposées lors de chaque rencontre, la moitié de nos membres sont issus des pays dits « du Sud », les publications sont nombreuses. LTT incarne par essence la diversité linguistique, travaillant sur une pluralité de langues, réunissant des chercheurs de nombreux pays non francophones.

Pour la première fois, nous nous réunissons dans un pays qui n'appartient pas à la Francophonie institutionnelle, répondant ainsi à un vœu cher à Mme Gendreau Massalou, ancienne rectrice de l'AUE, qui nous rejoindra bientôt pour suivre nos travaux. Nous connaissons tous les liens étroits qui unissent la lusophonie et la francophonie, deux espaces qui garantissent la diversité linguistique dans ce monde, deux espaces qui s'estiment mutuellement et dont la collaboration est parfaitement illustrée par cette rencontre. Nous n'avons, hélas, pas encore fait le pas de l'intercompréhension dans le cadre de nos journées scientifiques ; le français sera la langue de communication de cette rencontre, mais j'invite chaleureusement nos collègues lusophones à pratiquer sans complexe leur belle langue en sorte que dans les couloirs comme dans les coulisses, elle résonne dans nos oreilles et que tous nous puissions nous imprégner de ses sonorités originales.

Pour avoir aidé mon collègue Daniel Blampain à organiser les journées de Bruxelles en 2005, je mesure toute l'ampleur du travail abattu depuis de longs mois ici à Lisbonne pour nous accueillir dans les meilleures conditions. Au nom de l'ensemble du comité scientifique, je tiens à remercier vivement le professeur Teresa Lino pour son engagement permanent au sein du réseau ; hier comme membre de notre comité, aujourd'hui comme coordonnatrice d'une action de recherche en réseau et, avec Rute Costa et toute son équipe, comme cheville ouvrière de l'organisation de ces *Huitièmes Journées scientifiques*. Je voudrais particulièrement saluer leurs efforts pour lever les fonds nécessaires à l'organisation de cette rencontre et, dans le même temps, remercier tous nos mécènes : l'Agence universitaire de la Francophonie, la Délégation générale à la langue française et aux langues de France, la Fondation Calouste Gulbenkian et de nombreuses ambassades locales, à commencer par l'ambassade de France, qui nous fera l'honneur de nous recevoir ce soir dans son prestigieux *Palácio de Santos*.

La formation à la recherche est au cœur de nos préoccupations et nos journées scientifiques sont traditionnellement précédées de journées de formation, toujours très attendues. Soucieux de satisfaire au mieux les attentes, le comité avait pris le pari d'organiser cette année deux sessions distinctes. Malgré cela, une fois encore, nous avons dû jouer à bureau fermé. Il convient donc de féliciter très chaleureusement les formateurs qui ont si bien collaboré à la réussite de ces journées de formation : Rute Costa, Raquel Silva et Manuel Célio Conceição, d'une part, Max Silberztein et Slim Mesfar, d'autre part. Une météorologie favorable, la simplicité de l'accueil, la joie de vivre et une volonté commune d'échange resteront dans la mémoire de toutes celles et de tous ceux qui ont eu le privilège de suivre ces journées de formation.

Tout serait-il pour le mieux dans le meilleur des mondes possibles ? L'Agence universitaire de la Francophonie vit une période de refondation et la structure des programmes est repensée. Toute transition suscite les peurs et, dans ce contexte, celle de voir se terminer notre belle aventure est bien présente. Le comité scientifique du Réseau LTT – Marie-Claude L'Homme, Philippe Thoiron, Ibrahim Ben Mrad, Thierno Cissé, Franck Neveu, Xavier Blanco et moi-même – nous devons de mener à bon port les 670 chercheurs embarqués – souvent depuis longtemps – dans l'aventure LTT, vieille de vingt ans déjà. Vous entendrez bientôt parler d'une disparition des réseaux de chercheurs en sciences du langage et de leur renaissance au sein d'un « collectif de recherche » consacré la diversité linguistique, vraisemblablement organisé autour de trois axes : lexicologie et terminologie, traduction et interculturalité, observation et action sur les langues. Le Réseau LTT, qui continuera à exister comme association autonome – non pas contre vents et marées, mais avec la bénédiction de l'AUF –, participera bien entendu à ce collectif qu'il entend contribuer à fonder. Selon mes archives, nos premières journées scientifiques se sont déroulées à Fès en 1989. Vingt ans, n'est-ce pas le bel âge et n'avons-nous pas toute la vie devant nous ? Après Fès (1989), Mons (1991), Montréal (1993), Lyon (1995), Tunis (1997), Beyrouth (1999) et Bruxelles (2005), nous voici à Lisbonne. Je ne doute pas que nous nous rencontrerons encore en bien d'autres endroits : à Dakar, à Hanoi, à Recife... Appel est d'ores et déjà fait aux universités candidates pour organiser les neuvièmes, dixièmes et onzièmes journées ! D'ici là, il nous faudra, bien entendu compléter la prestigieuse collection des actes de nos rencontres, après sélection des meilleures contributions. Je me dois donc de vous souhaiter d'entendre, en ce lieu prestigieux, des communications pleines d'intérêt, même s'il vous faudra opérer des choix douloureux entre des sessions parallèles. Un programme détaillé vous permettra de vous retrouver parmi ce vaste ensemble, à condition que chacun suive bien la discipline d'un horaire très cadencé, dont chaque président de séance se fera le gardien vigilant.

Mesdames, Messieurs, je proclame donc ouvertes ces *Huitièmes Journées scientifiques* et vous remercie de votre fidélité à notre réseau. *Obrigado.*

Communications

Aleksandra Dziadkiewicz,
Centre L. Tesnière,
Université de Franche-Comté,
Besançon,
France,
aleksandra.d@hotmail.fr.

Vers un dictionnaire pragmatique français-polonais, polonais-français : quels critères pour le choix des entrées ?

Résumé

Cet article s'inscrit dans la phase préparatoire d'un projet de dictionnaire pragmatique informatisé français-polonais, polonais-français, ayant pour objet de recherche les énoncés lexicalisés, liés à des situations déterminées. Plus particulièrement, il s'agit d'étudier la question du choix des entrées du dictionnaire : d'abord, en examinant les critères de sélection décrits en lexicographie pragmatique, et ensuite, en les confrontant au contenu de deux dictionnaires pragmatiques de référence. Les résultats de cette analyse permettent de délimiter différents sous-ensembles d'expressions et de discuter leur statut, tout en adoptant des critères de sélection aussi bien formels que sémantiques.

Mots-clés : énoncés liés ; phraséologismes pragmatiques ; pragmatographie ; phraséologie.

Introduction

Dans cet article, nous proposons une analyse liminaire de différents types d'unités lexicales candidates à un projet de dictionnaire pragmatique informatisé français-polonais, polonais-français. Ce dictionnaire rassemblerait des formules routinières et structures linguistiques fixées par l'usage, qui fonctionnent comme des énoncés syntaxiquement autonomes, mais qui restent pragmatiquement liées à des situations déterminées, par exemple :

[1] On aura tout vu !

[2] Où est le drame ?

[3] J'ai vu pire.

[4] Si je m'attendais à ça !

[5] Mais tu te rends compte ?

[6] Et alors ?

[7] De quoi je me mêle ?

Les unités concernées figurent dans les typologies de phraséologismes sous des dénominations variées : entre autres, « énoncés liés » (Fónagy 1982), « phraséologismes pragmatiques » (Burger et al. 1982 ; Dziadkiewicz 2007 ; López Simó et Sierra Soriano 2005) ou encore « expressions d'illocutoire stéréotypé » (Schneider 1989). Nous utiliserons ici le terme d'énoncés liés, afin de nous situer dans la lignée des travaux d'Iván Fónagy, précurseur de l'idée même de dictionnaires listant les expressions de ce type.

L'une des questions fondamentales qui se posent dans un projet d'un dictionnaire pragmatique est celle du choix de ses entrées : quelles expressions faut-il y inclure ? Qu'englobe précisément la notion d'énoncé lié ? Où se trouvent ses limites ? Dans la première partie de l'article, nous exposerons brièvement les critères du choix des entrées en lexicographie pragmatique, présentés d'un point de vue théorique, pour ensuite, dans la seconde partie, examiner le contenu de deux dictionnaires pragmatiques de référence (Martins-Baltar 2002 ; Bidaud 2002), délimiter les différents sous-ensembles d'unités lexicales étudiées et motiver la décision de les retenir ou de les rejeter dans le cadre de notre projet, tout en signalant les difficultés et les hésitations qui ont pu surgir lors de cette analyse.

1 Critères du choix des entrées en lexicographie pragmatique

La lexicographie pragmatique, appelée également pragmatographie, est une « lexicographie non des mots, mais des formulations usuelles des actes de paroles, des phrases : elle a vocation à produire des dictionnaires d'énoncés, les énoncés étant définis par leurs conditions d'emploi et leurs fonctions, c'est-à-dire dans un cadre pragmatique » (Martins-Baltar 2009). Les fondements théoriques qui sous-tendent cette discipline sont basés dans une grande mesure sur l'analyse proposée par Iván Fónagy (1982, 1997). Selon l'auteur, les critères définitoires des énoncés liés, permettant de les délimiter par rapport aux énoncés libres, sont ceux de conventionnalité et de réduction sémantique : plus un énoncé est stéréotypé et récurrent dans certaines situations, plus il perd sa valeur sémantique au profit de sa fonction pragmatique et plus il a de chances d'être identifié comme un énoncé lié. Ainsi, dans l'exemple suivant

[8] – Elle sort tous les soirs ! – Et alors, où est le mal ?

la fonction pragmatique des énoncés « et alors » et « où est le mal ? » (objection à une critique, minimisation) relègue au second plan leur sens compositionnel. Cependant, comme nous le montrerons dans la deuxième partie de cet article, il n'est pas toujours aisé d'estimer un seuil de conventionnalité à partir duquel une construction récurrente serait retenue comme une entrée potentielle du dictionnaire, et la conventionnalité ne va pas obligatoirement de pair avec la réduction sémantique.

Quels sont d'autres critères de sélection et de description proposés par les auteurs de dictionnaires pragmatiques ? Si les recueils d'énoncés liés sont relativement peu nombreux, deux ouvrages offrent une base solide pour tout projet touchant à la pragmatographie française : dans une perspective monolingue, le *Dicomotus* de Michel Martins-Baltar (2002) et, en pragmatographie bilingue, *Structures figées de la conversation. Analyse contrastive français-italien* de Françoise Bidaud (2002).

1.1 *Dicomotus*

Le *Dicomotus* est une première tentative d'une fine description monolingue des « énoncés de motif usuels »¹, réunissant 585 entrées dont le mot-clé commence par la lettre A, et basée sur le modèle Motif-Réaction. Comme l'annonce l'auteur, « *Dicomotus* se présente [...] comme un dictionnaire sémantique de réalisations indirectes d'actes illocutoires « liées », « lexicalisées », montrant, en compréhension et en extension, que les actes de langage indirects se réalisent par des représentations de conditions d'action (notamment de motifs) en rapport fonctionnel avec leur motif d'énonciation » (Martins-Baltar 1997).

Sans entrer dans les détails de son organisation complexe, disons que le *Dicomotus* propose une caractérisation des énoncés de motif usuels par un motif d'énonciation (p. ex. : état de choses naturel ou humain, acte matériel, acte de parler), une fonction réactive par rapport à ce motif d'énonciation (fonctions cognitives ou axiologiques, d'opposition ou de participation, d'intervention ou d'évaluation) et un mode de réalisation de cette fonction (éléments lexicaux spécialisés, locutions ou implicites linguistiques).

1.2 Structures figées de la conversation. Analyse contrastive français-italien

Dans notre analyse, nous nous sommes également basée sur le recueil de Françoise Bidaud, qui est à nos yeux l'ouvrage le plus complet et le plus explicatif en ce qui concerne les dictionnaires pragmatiques bilingues mettant en cause le français. Dans ce dictionnaire nous trouvons, en effet, 1 057 entrées accompagnées de leurs définitions « traditionnelles » (le plus souvent inspirées ou copiées de dictionnaires généraux de référence, comme *Larousse* ou *Robert*), mais aussi des étiquettes sémantiques, des paraphrases, des explications supplémentaires concernant l'emploi des formules, ainsi que des attestations de ces expressions en contexte.

La caractérisation générale des entrées, donnée par Bidaud, reste dans un esprit similaire au modèle Motif-Réaction de Martins-Baltar : « Les énoncés relevés sont des répliques, des commentaires qui s'insèrent souvent – ou peuvent s'insérer – dans un échange et n'acquièrent de sens qu'à l'intérieur d'un contexte. [...] Il peut même s'agir d'un commentaire à une attitude, d'une réaction à du non-verbal » (Bidaud 2002 : 5).

1 Les énoncés de motif usuels étant un sous-ensemble des énoncés liés (Martins-Baltar 1997).

Cependant, ce qui mérite une attention plus particulière dans la description de Françoise Bidaud, ce sont les étiquettes sémantiques qu'elle a attribuées à toutes les entrées : celles-là signalent « le (ou les) sentiment(s) que l'expression permet de manifester. Cela implique une analyse préalable du contenu sémantique de chaque expression, une interprétation, résumée ensuite sous forme de concept » (2002 : 13). Ainsi, les étiquettes sémantiques permettent, premièrement, d'accéder rapidement et de façon claire au concept-clé se trouvant à la base de chaque expression, deuxièmement, d'établir un classement sémantique compréhensible par tous les apprenants de français, et non seulement par les italophones, et, enfin, de préciser les éléments sémantiques que les expressions équivalentes dans les deux langues ont en commun. Or, l'objectif de notre dictionnaire étant non seulement de proposer des équivalences en français et en polonais, mais aussi d'expliquer quels éléments sémantiques et situationnels permettent d'établir ces équivalences, la méthode de description au moyen d'étiquettes sémantiques paraît entièrement adéquate à notre projet, à condition toutefois de la développer, en étoffant les étiquettes de paramètres plus variés². D'autre part, un examen sommaire des appellations figurant dans les étiquettes sémantiques montre que celles-ci désignent en pratique bien plus que des sentiments : les concepts proposés sont en effet de nature très variée et laissent présager de la diversité des « structures figées de la conversation ».

[9] Accomplissement : C'est chose faite !

[10] Coïncidence : Ça tombe bien ! Tiens !

[11] Conséquence : Ça va (ira) loin ! Comme quoi...

[12] Incapacité : Comme un pied !

[13] Opinion : Je me suis laissé dire que... ; Il y a le pour et le contre !

[14] Conseil : Laisse tomber ! Tâche moyen de...

[15] Politesse : Prenez la peine de... ; Qu'est-ce que tu deviens ?

[16] Silence : Un ange passe.

[17] Incise : Soit dit en passant...

[18] Attaque de discours : Quoi de neuf ?

[19] Précaution oratoire : Si j'ose dire.

2 Que trouve-t-on dans les dictionnaires pragmatiques ?

Cette première appréciation de la diversité des entrées analysées nous amène à la deuxième partie de cet article, où nous tenterons de répondre aux questions suivantes : que trouve-t-on dans les dictionnaires pragmatiques ? Quels types

² Nous avons déjà utilisé cette méthode dans Dziadkiewicz (2007), pour décrire un groupe de phraséologismes pragmatiques français et polonais, réalisant six actes de langages marqués au niveau de la politesse : les étiquettes sémantico-pragmatiques composées de cinq paramètres nous ont servi de pivot permettant d'établir des équivalences entre les expressions des deux langues.

d'expressions peut-on y distinguer? Dans le cadre du projet d'un dictionnaire pragmatique bilingue, comment motiver la décision de retenir ou d'écarter ces différents types? Et enfin, où situer la frontière entre, d'une part, les énoncés liés et les locutions et, d'autre part, entre les énoncés liés et les constructions libres?

2.1 Énoncés présentant des anomalies syntaxiques ou sémantiques

Le premier sous-ensemble d'expressions réunit toutes les caractéristiques permettant de les retenir comme entrées du dictionnaire: il s'agit d'énoncés autonomes, de forme fixe, présentant des anomalies syntaxiques ou sémantiques.

[20] Les bêtises qu'on peut trouver dans ce bouquin, *je vous dis pas!*³ (Omission de COD, réduction sémantique)

[21] – Vous avez les calculs de correction? – *À quoi bon. C'est trop compliqué.* (Expression sémantiquement opaque, structure syntaxique particulière)

Le figement pragmatique de certains énoncés peut être vérifié au moyen de tests: par exemple, on observe dans certaines situations une forte préférence pour une forme donnée, au point de considérer les formes synonymiques comme pragmatiquement et sémantiquement incorrectes.

[22] – Je ne suis absolument pas satisfaite de sa conduite. – Ah ben, *tu m'étonnes!* (mais non: *Tu m'épates; Tu me surprends*).

Un autre test consiste à soumettre l'énoncé candidat à des transformations: si certaines d'entre elles sont inacceptables ou si elles modifient l'interprétation de l'énoncé, on ne peut pas parler d'énoncé libre. Ainsi, l'énoncé lié dans l'exemple ci-dessus exprime l'accord, l'adhésion à une opinion: valeur qui se perd dès que l'on change de personne, de mode, ou de temps:

[23] Il m'étonne

[24] Étonner qqn

[25] Tu m'as étonné

[26] Ça m'étonnerait (ce dernier exemple est un autre énoncé lié qui, lui, exprime l'incrédulité).

2.2 Énoncés à trous

Les expressions du deuxième groupe peuvent soulever des doutes quant à leur forme, puisqu'elles ne constituent pas des énoncés clos: elles combinent en effet des éléments fixes avec des éléments variables:

[27] Prenez la peine de... [vous asseoir, entrer...]

³ Les énoncés analysés sont soulignés lorsqu'ils sont insérés dans un contexte plus long.

[28] L'autobus devrait passer toutes les dix minutes et ça fait déjà une demi-heure qu'on attend! – *Tu vois pas qu'il y ait grève?*⁴

Ces expressions devraient être retenues, dans la mesure où elles se présentent comme l'élément-clé de l'énoncé et sont porteuses de valeurs illocutoires spécifiques (ici: invitation à une action; expression de supposition et de souhait). Par ailleurs, les transformations enlèvent ces valeurs illocutoires et entraînent d'autres choix lexicaux en traduction, ce qui est un autre argument en faveur de ce groupe:

[29] Se donner la peine – *zadawać sobie trud, starać się* (le sens est de: faire l'effort de)

[30] Les passants, maintenant, ne se donnaient plus la peine de cacher leur curiosité – *Przechodnie nie starali się już ukrywać swojej ciekawości.*

[31] Monsieur, donnez-vous la peine d'entrer – *Proszę, niech pan zechce wejść.*

2.3 Énoncés formés d'un syntagme ou d'un mot

Le groupe suivant rassemble les expressions formées d'un syntagme voire d'un seul mot: de ce fait, il pourrait être sujet à discussion, la polylexicalité étant souvent admise comme l'un des critères définitoires pour qu'on puisse parler de figement et d'unités phraséologiques. Cependant, si l'on élargit la notion de figement à celle de figement pragmatique, on s'aperçoit qu'un syntagme ou un seul mot, dans une situation d'énonciation, peut fonctionner comme un énoncé lié:

[32] – Où est-il? – Mais à son travail, *cette idée!*

[33] – Et si on allait au cinéma? – *Ma foi!* Je ne dis pas non.

[34] Allez, va-t'en, je ne veux plus te voir! *De l'air...*

[35] – Elle a eu son bac? – *À l'aise! Les doigts dans le nez!*

[36] (Bien) entendu!

[37] Pouce!

[38] Touché!

Ce type de figement doit être d'autant plus pris en compte dans un contexte interlinguistique, puisqu'un énoncé lié formé d'un seul mot ne saurait être rendu littéralement dans une autre langue sans produire une traduction inintelligible.

2.4 Énoncés liés vs locutions

La discussion autour des critères formels de sélection des entrées touche également à une question qui mérite d'être examinée à part, celle de la différence entre les énoncés liés et les locutions. En effet, dans les deux dictionnaires consultés, on

⁴ Bidaud (2002 : 285) en donne la paraphrase suivante: Sens: « et s'il se passait ceci! » sous-entendu: « j'espère que ça n'arrivera pas. »

trouve des énoncés qui ne sont en réalité que des formes actualisées de locutions verbales, adverbiales, etc. :

- [39] – Les voisins lui ont demandé de garder leur petite fille. – Alors, elle *est à son affaire!* Elle adore les enfants. (= être à son affaire)
- [40] – Elle voudrait prendre ses vacances au mois de juin ? – Oui, mais son patron *ne l'entend pas de cette oreille!* (= ne pas l'entendre de cette oreille)
- [41] – Il a trouvé que les enfants de Pierre étaient mal élevés et il *ne le lui a pas envoyé dire!* (= ne pas l'envoyer dire à qqn)
- [42] Je vais faire les courses et *par la même occasion* je passerai à la poste.
- [43] – C'est encore loin ? – *À vue de nez*, encore un kilomètre.
- [44] *Jamais, au grand jamais*, je ne me suis occupé de marché noir.
- [45] En mon âme et conscience

Nous ne retenons pas ces locutions dans le cadre du dictionnaire d'énoncés liés, même s'il faut reconnaître pleinement la nécessité de les décrire en contexte, et de préférence dans des séquences dialogiques. D'une part, on ne peut parler de fixité formelle de ces énoncés, puisque les locutions verbales ci-dessus peuvent se conjuguer à toutes les personnes, temps et modes, sans que leur sens soit modifié. D'autre part, la décision de rejet se trouve corroborée par la confrontation des exemples cités avec les critères sémantiques annoncés dans le premier paragraphe : en effet, les énoncés ci-dessus évoquent des concepts qui ne participent pas à la situation d'énonciation. Ainsi, les étiquettes sémantiques des énoncés 39 à 41 correspondent, respectivement, à « Satisfaction/Aptitude », « Désaccord » et à « Brutalité », mais dans aucun des cas il ne s'agit de sentiments ou d'attitudes des interlocuteurs : les concepts se réfèrent aux personnes externes à la situation d'énonciation.

Une exception parmi les locutions est toutefois à signaler : certaines locutions présentent des caractéristiques illocutoires leur permettant de fonctionner comme des actes d'affirmation ou de négation, accompagnées d'autres valeurs sémantiques : facilité, promesse, engagement, etc. Ces unités seraient donc retenues.

- [46] Elle a eu son bac ? – *À l'aise!* Les doigts dans le nez !
- [47] Alors dites, vous en pincez pour elle ? – *Mais jamais de la vie*, vous êtes folle !
- [48] Alors, à ce soir ? – *Oh, sans faute!*

2.5 Phrases situationnelles

Les phrases situationnelles, formant un autre sous-ensemble d'énoncés liés, présentent la particularité de fonctionner comme des réactions relatives à des situations bien déterminées, c'est-à-dire impliquant des conditions d'emploi précises :

- [49] Un ange passe. (Silence)

[50] Tu as perdu ta langue ? (Silence)

[51] Ton père n'est pas vitrier ! (Protestation)

[52] Tu veux ma photo ? (Insolence)

On peut remarquer une certaine incohérence ou insuffisance des étiquettes sémantiques attribuées par Bidaud : dans les énoncés 49 et 50, elles désignent l'une des conditions de leur emploi, dans 51, il s'agit de l'acte de langage exprimé, et dans 52, de l'attitude du locuteur. D'autre part, il serait souhaitable de développer les étiquettes, en y intégrant des éléments de définitions traditionnelles, en proposant une décomposition sémantique et une schématisation des conditions d'énonciation : d'abord pour pouvoir faire la distinction entre les énoncés comme 49 et 50, et, ensuite, pour rendre compte de la nature différente de ces phrases situationnelles par rapport aux énoncés exprimant des sentiments, liés à des situations plus générales.

2.6 Proverbes

Le groupe suivant est constitué de proverbes, dont voici quelques exemples :

[53] Deux avis valent mieux qu'un.

[54] Les absents ont toujours tort.

[55] Un homme averti en vaut deux.

[56] Il n'y a pas de petites économies !

Les proverbes ont plusieurs propriétés en commun avec les énoncés liés : ce sont des énoncés clos et autonomes, de forme fixe, déclenchés par des situations déterminées. Cependant, deux raisons font qu'il est préférable de les lister dans un recueil de proverbes, plutôt que de les intégrer dans un dictionnaire d'énoncés liés. D'abord, leur grand nombre, qui nécessite une analyse à part ; et ensuite leur nature différente : les proverbes sont des phrases génériques, sans rapport déictique avec la situation d'énonciation, contrairement aux phrases épisodiques, qui elles constituent des assertions sur des faits particuliers :

[57] C'est l'hôpital qui se moque de la charité.

2.7 Expressions servant à organiser ou à modaliser le discours

Les expressions de ce groupe figurent dans le dictionnaire de Bidaud accompagnées d'étiquettes qui ne font référence ni aux sentiments ni aux attitudes : les noms employés évoquent l'organisation et la modalisation du discours :

[58] Encore faut-il que... (Condition)

[59] A tout prendre ; Tout compte fait (Conclusion)

[60] À proprement parler (Précision)

- [61] N'empêche... (Objection/Digression)
- [62] A ce compte-là... (Conséquence)
- [63] À ce que j'en sais (Opinion/Information Partielle)
- [64] Si ça se trouve... (Possibilité)

Ces expressions, si elles peuvent avoir leur place au sein des « structures figées de la conversation », sont en revanche à exclure d'un dictionnaire d'énoncés liés, puisqu'on ne saurait parler ici de liage entre l'énoncé et la situation.

2.8 Critères sémantiques

Nous avons évoqué, dans les points précédents, les critères sémantiques à l'appui de nos choix concernant les différents sous-ensembles délimités. Pour résumer, voici les catégories d'étiquettes que nous avons délimitées et adoptées comme critères de sélection sémantiques :

– sentiments :

[65] On en apprend tous les jours! (Étonnement);

– actes de langage :

[66] Écrase! (Repousser)

[67] Je te le fais pas dire! (Accord);

– attitudes :

[68] Il faut faire avec! (Résignation)

[69] J'y suis, j'y reste! (Opiniâtreté);

– commentaires portant sur son propre discours :

[70] Je disais ça comme ça (Minimisation)

[71] Ne me faites pas dire ce que je n'ai pas dit! (Rétractation);

– commentaires portant sur la situation :

[72] Ça va (ira) loin! (Conséquence)

[73] C'est pas gagné (Difficulté)

[74] Ça promet (Perspective pessimiste).

2.9 Énoncés liés vs énoncés libres

Le dernier point concerne la question de la frontière entre les énoncés liés et les énoncés libres. Nous avons vu, dans le premier paragraphe, que les critères défini-
toires des énoncés liés étaient ceux de conventionnalité et de réduction sémantique. Or, certaines expressions du *Dicomotus*, bien que récurrentes, ne semblent pas répondre à ces critères. C'est le cas, par exemple, de la structure « C'est + Adjectif »,

qui soulève un doute quant à sa fixité formelle, d'autant plus que les entrées sont accompagnées de variantes où seul l'adjectif, voire seule la racine, rapproche la variante de l'entrée principale :

[75] C'est affreux.

[76] C'est aberrant. Il est aberrant de ne pas [faire].

[77] C'est abominable. Ça me paraît absolument abominable.

[78] C'est assommant. Les gens l'assommaient. Ça l'assomme.

Il faut cependant signaler les cas, chez Bidaud, où la même structure présente un changement sémantique et qui de ce fait seraient à retenir :

[79] C'est gai. (Contrariété)

[80] C'est bon. (Accord, Impatience)

D'autres énoncés du *Dicomotus* répondent au critère de conventionnalité, mais pas à celui de réduction sémantique :

[81] Il faut que je te fasse un aveu.

[82] Il fallait le faire avant.

[83] Pas d'utilisation prolongée sans avis médical.

Par conséquent, nous avons choisi de les écarter de notre projet, afin de fixer des limites à ce travail. Cependant, on ne saurait nier l'importance et l'utilité de répertorier ce type d'énoncés usuels, notamment à l'usage de la didactique et de la traduction.

Conclusion

Nous avons vu, tout au long de cette analyse, que les variations dans la nature des étiquettes sémantiques étaient révélatrices de la complexité des sens et des conditions d'emploi des expressions étudiées. L'observation du contenu des deux dictionnaires consultés nous a permis de délimiter des sous-ensembles d'expressions et de retenir un certain nombre d'entre eux, en adoptant des critères de sélection aussi bien formels que sémantiques. Enfin, la présente étude a révélé la nécessité d'adapter la forme et les paramètres des étiquettes sémantico-pragmatiques aux particularités de chaque type d'expressions.

Bibliographie

- Bidaud (F.), 2002 : *Structures figées de la conversation. Analyse contrastive français-italien*, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Oxford, Wien, Peter Lang.
- Burger (H.) et al., 1982 : *Handbuch der Phraseologie*, Berlin, New York, de Gruyter.

- Dziadkiewicz (A.), 2007 : *Vers une reconnaissance et une traduction automatique de phraséologismes pragmatiques. Application du français vers le polonais*, thèse de doctorat (confidentielle), Université de Franche-Comté, Besançon.
- Fónagy (I.), 1982 : *Situation et signification*, Amsterdam, Benjamins.
- Fónagy (I.), 1997 : « Figement et changements sémantiques », dans Martins-Baltar (M.), éd., *La locution entre langue et usages*, Fontenay/Saint-Cloud, ENS éditions, p. 131-164.
- López Simó (M.) et Sierra Soriano (A.), 2005 : « Fraseologismos pragmáticos reactivos: análisis contrastivo francés español », dans *Actes du Congrès international Première Rencontre hispano-française de Chercheurs (SHF-APFUE)*, 29 novembre – 2 décembre 2005, Séville, p. 873-884.
- Martins-Baltar (M.), 1997 : « De l'énoncé lié à l'énoncé de motif usuel: le projet de dictionnaire *Dicomotus* », dans Perrot (J.), éd., *Polyphonie pour Ivan Fónagy*, Paris, L'Harmattan, p. 323-341.
- Martins-Baltar (M.), 2000 : « Construire un dictionnaire d'énoncés (unilingue comme plurilingue) : à propos de *Dicomotus* », dans Szende (Th.), éd., *Approches contrastives en lexicographie bilingue*, Paris, Champion, p. 301-317.
- Martins-Baltar (M.), 2002 : *Dicomotus*, dernière mise à jour en 2002, <http://michel.martinsbaltar.neuf.fr/ds/Dicomotus.html>
- Martins-Baltar (M.), 2006 : « Comment construire une nomenclature en pragmatographie ? » dans Szende (Th.), éd., *Le français dans les dictionnaires bilingues*, Paris, Champion, p. 193-303.
- Martins-Baltar (M.), 2009 : *Pragmalex: matériaux de pragmatographie et de lexicographie*, dernière mise à jour le 2 octobre 2009, <http://michel.martinsbaltar.neuf.fr/index.html>
- Schneider (F.), 1989 : *Comment décrire les actes de langage? De la linguistique pragmatique à la lexicographie: « La belle affaire! » et « Tu m'en diras tant! »*, Tübingen, Niemeyer.

Mathieu Mangeot,
Laboratoire GETALP-LIG 385 rue de la Bibliothèque BP 53,
F-38041 GRENOBLE CEDEX 9,
France,
Mathieu.Mangeot@imag.fr.

Hong-Thai Nguyen,
Laboratoire LLS Université de Savoie BP 1104,
F-73011 CHAMBÉRY CEDEX,
France,
Hong-Thai.nguyen@imag.fr.

Projet MotÀMot : élaboration d'un système lexical multilingue par le biais de dictionnaires bilingues

Résumé

Le projet MotÀMot d'élaboration d'un système lexical multilingue est ciblé sur des langues d'Asie du Sud-Est, en particulier le vietnamien et le khmer. La macrostructure est une structure pivot avec un volume pour chaque langue autour d'un volume pivot reliant les sens de mot de chaque langue entre eux. La microstructure est basée sur la lexicographie explicative et combinatoire. Les contributions se feront en ligne, sur la plate-forme Jibiki, par des communautés de bénévoles constituées autour de jeux sérieux lexicaux. Chaque donnée se verra attribuer un niveau de qualité, de même pour chaque contributeur.

Mots-clés : lexicographie multilingue ; langues peu dotées ; projet contributif ; projet MotÀMot.

1 Introduction et contexte

1.1 Informatisation des langues peu dotées

Les enjeux économiques liés aux techniques du traitement de l'information sont très importants. Le développement de telles technologies est un atout majeur pour des pays en voie de développement comme le Cambodge et le Laos, ou émergents comme le Vietnam, la Malaisie et la Thaïlande.

Comme l'indique V. Berment dans sa thèse (Berment 2004), « le développement des ordinateurs personnels et celui des réseaux font aujourd'hui de l'informatique un instrument pour écrire et communiquer au même titre que le papier et

l'imprimerie l'étaient auparavant. Traitements de texte, courriers électroniques, voire des systèmes plus avancés comme la dictée ou la synthèse vocale sont des outils largement répandus. L'idée s'impose alors qu'aux moyens traditionnels doivent s'ajouter les outils informatiques appropriés sans lesquels les buts visés ne peuvent plus être atteints ». L'informatisation d'une langue occupe ainsi une place essentielle dans ce vaste contexte.

Cependant, parmi les 6000 langues parlées dans le monde, seul un tout petit nombre d'entre elles atteint un « niveau d'informatisation » satisfaisant. Pour évaluer de manière quantitative le degré d'informatisation d'une langue, il propose le protocole suivant : à chaque service ou ressource, un groupe d'utilisateurs représentatifs des locuteurs de la langue attribue un niveau de criticité C_k et une note N_k , la moyenne pondérée des notes – appelée indice – reflétant leur satisfaction globale. Une langue mal ou peu dotée peut ainsi être définie comme une langue dont l'indice n'atteint pas 10/20. À titre d'exemple, on peut présenter dans les deux figures ci-dessous une évaluation du niveau d'informatisation obtenu ainsi pour la langue khmère, parlée au Cambodge, et pour la langue vietnamienne. La figure 1 représente l'évaluation du niveau d'informatisation pour le khmer et la figure 2 pour le vietnamien.

| Services / ressources | Criticité (/10) | Note (/20) | Note pondérée (Criticité x Note) |
|--------------------------------------|-----------------|------------|----------------------------------|
| Traitement du texte | | | |
| Saisie simple | 10 | 16 | 160 |
| Visualisation / impression | 10 | 14 | 140 |
| Recherche et remplacement | 8 | 12 | 48 |
| Sélection du texte | 6 | 12 | 72 |
| Tri lexicographique | 5 | 0 | 0 |
| Correction orthographique | 2 | 0 | 0 |
| Correction grammaticale | 0 | 0 | 0 |
| Correction stylistique | 0 | 0 | 0 |
| Traitement de l'oral | | | |
| Synthèse vocale | 5 | 0 | 0 |
| Reconnaissance de la parole | 5 | 0 | 0 |
| Traduction | | | |
| Traduction automatisée | 8 | 4 | 32 |
| ROC | | | |
| Reconnaissance optique de caractères | 9 | 0 | 0 |
| Ressources | | | |
| Dictionnaire bilingue | 10 | 4 | 40 |
| Dictionnaire d'usage | 10 | 0 | 0 |
| Total | | | 540 / 1760 |
| Moyenne | | | 6,2 / 20 |

Figure 1

| Services / ressources | Criticité (/10) | Note (/20) | Note pondérée (Criticité x Note) |
|--------------------------------------|-----------------|------------|----------------------------------|
| Traitement du texte | | | |
| Saisie simple | 10 | 16 | 160 |
| Visualisation / impression | 10 | 16 | 160 |
| Recherche et remplacement | 8 | 17 | 136 |
| Sélection du texte | 6 | 17 | 102 |
| Tri lexicographique | 5 | 6 | 30 |
| Correction orthographique | 2 | 6 | 12 |
| Correction grammaticale | 0 | 0 | 0 |
| Correction stylistique | 0 | 0 | 0 |
| Traitement de l'oral | | | |
| Synthèse vocale | 5 | 0 | 0 |
| Reconnaissance de la parole | 5 | 0 | 0 |
| Traduction | | | |
| Traduction automatisée | 8 | 6 | 48 |
| ROC | | | |
| Reconnaissance optique de caractères | 9 | 12 | 108 |
| Ressources | | | |
| Dictionnaire bilingue | 10 | 13 | 130 |
| Dictionnaire d'usage | 10 | 0 | 0 |
| Total | | | 886 / 1760 |
| Moyenne | | | 10 / 20 |

Figure 2

Comme nous pouvons le voir sur ces figures, les services liés au traitement de l'oral (ou technologies vocales, c'est-à-dire la synthèse vocale et la reconnaissance de la parole) sont inexistantes pour ces deux langues. C'est aussi le cas pour une majorité de langues dans le monde, dont certaines sont parlées par plusieurs dizaines de millions de locuteurs (par exemple¹ le bengali : 189 millions, le tamoul : 63 millions), y compris au sein de l'Europe des 25 (lituanien, letton, polonais) !

1.2 État des lieux de la lexicographie bilingue

La principale difficulté actuelle de la lexicographie bilingue réside dans les coûts prohibitifs de construction de grandes quantités de données. Par exemple, le projet *Electronic Dictionary Research* (EDR) dont le but était de construire un dictionnaire japonais-anglais a nécessité plus de 1 200 personnes/année de travail. Son prix de vente, de 84 000 € environ, est très inférieur aux coûts réels de la construction, coûts qui ne seront probablement jamais rentabilisés.

De toute façon, ces coûts sont trop élevés pour un particulier. De ce fait, seules des institutions peuvent l'acquérir. D'autre part, les données fournies à ce prix ne sont utilisables que par certains systèmes de traduction automatique fondés sur des techniques particulières.

Face à ces coûts difficilement gérables, les maisons d'édition finissent par vivre sur leurs acquis et ne proposent principalement que de nouvelles éditions de dictionnaires existants. Rares sont les éditeurs à avoir le courage de se lancer dans la réalisation d'un nouveau dictionnaire bilingue de qualité en partant de zéro.

¹ www.populationdata.net

D'autre part, même dans les dictionnaires les plus complets, on constate quasiment toujours un manque d'informations, en particulier concernant les collocations. Les rares ressources qui en tiennent compte ne le font pas de manière systématique.

Malgré l'arrivée d'Internet, il existe à l'heure actuelle peu de ressources lexicales de bonne qualité disponibles gratuitement en ligne. La plupart sont en fait des lexiques bilingues faits par des bénévoles non spécialistes en lexicographie.

La lexicographie multilingue en tant que telle n'en est en fait qu'à ses débuts. En effet, il n'existe pas vraiment de moyen d'imprimer un vrai « dictionnaire multilingue ». Il est par contre tout à fait possible de trouver des bases terminologiques multilingues (comme IATE) ou bien, à la rigueur, des petits lexiques ou livres de phrases multilingues.

Il n'a par ailleurs pas été encore suffisamment prouvé que la réutilisation d'un dictionnaire d'un couple de langue A → langue B pour en construire deux autres langue B → langue C et langue A → langue C était réellement avantageuse. C'est donc ce à quoi nous voulons nous attacher avec ce présent projet.

2 Objectifs du projet

Avec l'objectif général de participer à l'informatisation des langues peu dotées, ce projet consiste à élaborer un système lexical multilingue en construisant simultanément plusieurs dictionnaires bilingues partageant au moins une langue entre eux. La construction des dictionnaires bilingues se fera en ligne sur un site de type « Papillon » construit sur la plate-forme Jibiki selon une méthodologie de travail collaboratif et bénévole, inspirée du projet Wikipédia.

Les liens bilingues créés lors de la rédaction des articles sont utilisés d'une part pour générer des liens bilingues inverses, et d'autre part pour créer de nouveaux liens interlingues.

Ces dictionnaires seront aussi disponibles sous la forme d'une version multimédia, avec une interface conviviale et ergonomique, dont les résultats seront accessibles par deux médias : langue (texte) et parole.

Pour certaines langues (que nous traiterons comme exemples), des modules de synthèse seront rajoutés pour permettre de participer à l'apprentissage de la langue et de donner des exemples sonores de type « mots isolés » ou « livre de courtes phrases » aux apprenants.

- Les trois objectifs principaux de ce projet sont donc le lancement d'une dynamique de contribution autour de la construction de chaque dictionnaire bilingue en présence. Le succès de Wikipédia montre que cela est possible, à condition d'avoir des outils simples et faciles à utiliser ; le passage à grande échelle d'expériences de laboratoire telles que la base DiCo (Mel'čuk et Polguère 2006) ou le

système parax (Blanc 1996); et enfin l'élaboration d'un terrain d'expérimentation pour la validation de plusieurs hypothèses formulées dans de précédents travaux:

- Bijektivité des liens bilingues et transitivité des liens interlingues;
- Contribution massive sur le Web;
- Construction d'un système lexical multilingue (Polguère 2006).

3 Avancées dans la construction de ressources en ligne

3.1 Sur l'architecture des ressources multilingues : le projet Papillon

Une solution parfaite, le graal des ressources lexicales, serait une base de données lexicales multilingue à structure pivot, de bonne qualité et large couverture avec des entrées monolingues riches et des liens interlingues, utilisable aussi bien par des humains que par des machines, éditable en ligne et disponible gratuitement. Nous avons lancé en 2000 le projet Papillon² de construction d'une base multilingue pour tenter d'avancer dans cette direction.

La macrostructure est constituée d'un volume monolingue pour chaque langue et d'un volume pivot au centre (voir figure 3). Lorsqu'un nouvel article dans une langue A est ajouté, il doit être relié au volume interlingue. Ces liens sont créés soit en réutilisant des dictionnaires bilingues existants langue A → langue B, soit en les ajoutant manuellement à partir d'une traduction. Le lien langue A → langue B devient langue A → pivot → langue B. Si l'article langue B est déjà relié à un autre article langue C, alors l'article langue A bénéficiera lui aussi de ces liens.

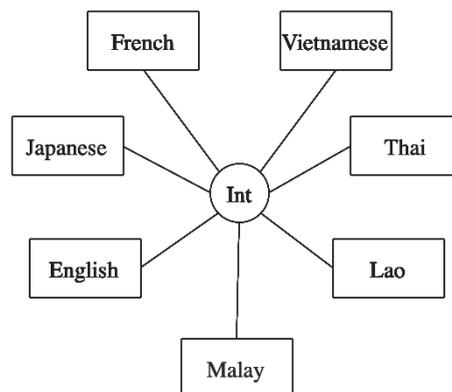


Figure 3

La microstructure des articles monolingues est riche et très détaillée. Elle est basée sur la structure utilisée pour la base lexicale DiCo (Polguère 2000) de l'OLST, Université de Montréal. La méthodologie d'encodage est directement empruntée à la

² www.papillon-dictionary.org

lexicologie explicative et combinatoire, partie de la théorie Sens-Texte. Cette théorie donne les informations nécessaires pour passer d'un sens à ses réalisations dans une langue donnée. La microstructure des dictionnaires est donc indépendante des langues et peut être théoriquement utilisable par des humains et par des machines.

Chaque article ou unité lexicale est basé sur le sens de mot ou lexie. Il est constitué d'un nom, des propriétés grammaticales, d'une formule sémantique qui peut être vue comme une définition formelle – dans le cas d'une lexie, prédicative, la formule décrit le prédicat et ses arguments et on trouve aussi le régime qui décrit la réalisation syntaxique des arguments –, puis d'une liste de fonctions lexico-sémantiques – il y a 56 fonctions de base applicables à toute langue et pouvant se combiner entre elles –, d'une liste d'exemples et d'une liste d'expressions idiomatiques.

Les spécifications du projet Papillon sont directement inspirées de ce fameux graal. Mais comme tout projet ambitieux, il ne peut être achevé d'un coup. Avec le temps, le projet Papillon est devenu une sorte de cadre ou métaprojet (Mangeot 2006) avec plusieurs projets dérivés, chacun correspondant à un aspect particulier de notre but initial. Comme nous le détaillerons par la suite, les aspects outils et systèmes sont couverts par le projet Jibiki et la collecte de données par le projet JeuxDeMots.

3.2 Sur les aspects contributifs : les projets Wikipedia et Wiktionary

L'encyclopédie en ligne contributive Wikipédia³ a rencontré un gros succès incontestable. On aurait pu s'attendre à une réussite comparable pour son petit frère Wiktionary⁴ mais le succès n'est pas encore au rendez-vous (1,5 million d'entrées pour le français et seulement 44 000 pour le japonais). Wiktionaryz⁵, qui prétendait parer aux défauts de Wiktionary n'a pas non plus eu l'effet escompté. Wiktionary n'est de toute façon pas un véritable dictionnaire bilingue même si on trouve quelques liens de traduction (il manque entre autres les indications de contexte de traduction).

Une hypothèse pour expliquer ce problème est celle de la motivation. En effet, lorsqu'une personne contribue à un article de Wikipédia, elle est récompensée par la renommée. Elle sera ensuite reconnue comme un expert dans son domaine. Cela n'est pas possible avec un dictionnaire. Les contributions portent sur des petites parties d'informations très ciblées et sont de ce fait anonymes. D'autre part, il y a un aspect technique lié à la structure. Un article d'encyclopédie a une structure plus ou moins libre tandis qu'une entrée de dictionnaire doit suivre une structure très précise (mot-vedette, informations grammaticales, blocs sémantiques, bloc de traduction, blocs d'exemples, etc.). Il n'est donc pas possible de réutiliser une plate-forme wiki pour construire un dictionnaire avec une structure bien définie.

Une fois acceptée l'idée que rédiger des entrées de dictionnaire n'est pas aussi plaisant que travailler sur un article de Wikipédia, il faut trouver des solutions pour

³ fr.wikipedia.org

⁴ fr.wiktionary.org

⁵ www.wiktionaryz.com

motiver une communauté de bénévoles à contribuer à un dictionnaire. Les jeux sérieux lexicaux constituent une première piste. Il faut aussi mettre en valeur les contributeurs à travers par exemple un tableau des meilleurs contributeurs du mois. Et enfin, l'exploitation de réseaux communautaires tels que Facebook devrait aussi apporter de l'eau au moulin.

3.3 Sur la collecte de données via des jeux sérieux : le projet JeuxDeMots

JeuxDeMots (Lafourcade et Joubert 2008) est une tentative de réponse au problème précédent. Ce projet a pour but de construire un réseau lexical riche et évolutif, qui peut être comparé à un certain degré à la fameuse base WordNet (Miller *et al.* 1990). Le principe est le suivant : une partie nécessite deux joueurs. Lorsqu'un joueur A débute une partie, une consigne concernant un type de compétence (synonymes, contraires, domaines) est affichée, ainsi qu'un mot *M*. tiré aléatoirement dans une base de mots. Le joueur A a alors un temps limité pour répondre en donnant des propositions répondant, selon lui, à la consigne appliquée au mot *M*. Ce même mot, avec cette même consigne, est proposé à un autre joueur B ; le processus est identique.

Les deux demi-parties, celle du joueur A et celle du joueur B, ne sont pas simultanées, mais asynchrones. Pour toute réponse commune dans les propositions de A et B, ces deux joueurs gagnent un certain nombre de points. La structure du réseau lexical que nous cherchons ainsi à obtenir s'appuie sur les notions de nœuds et de relations entre nœuds pour construire un réseau du type de (Polguère, 2006). Chaque nœud du réseau est constitué d'une unité lexicale (terme ou expression) regroupant toutes ses lexies et les relations entre nœuds traduisent des fonctions lexicales. La première version du jeu pour le français a été lancée en juillet 2007. Il existe aussi des versions anglaises, arabes, japonaises, vietnamiennes, et thaï. Elles sont disponibles sur le Web⁶.

3.4 Sur les aspects techniques : la plate-forme Jibiki

Jibiki (Mangeot *et al.* 2004) est une plate-forme générique en ligne pour manipuler des ressources lexicales avec gestion d'utilisateurs et groupes, consultation de ressources hétérogènes et édition générique d'articles de dictionnaires. C'est un site Web communautaire développé au départ pour le projet Papillon. La plate-forme est programmée entièrement en Java, basée sur l'environnement « Enhydra ». Toutes les données sont stockées au format XML dans une base de données (Postgres). Ce site Web propose principalement deux services : une interface unifiée permettant d'accéder simultanément à de nombreuses ressources hétérogènes (monolingues, dictionnaires bilingues, bases multilingues, etc.) et une interface d'édition spécifique pour contribuer directement aux dictionnaires disponibles sur la plate-forme.

L'éditeur est basé sur un modèle d'interface HTML instancié avec l'article que l'on veut éditer. Le modèle peut être généré automatiquement depuis une description

6 jeuxdemots.org

de la structure de l'entrée à l'aide d'un schéma XML. Il peut être modifié ensuite pour améliorer le rendu à l'écran. La seule information nécessaire à l'édition d'un article de dictionnaire est donc le schéma XML représentant la structure de cette entrée. De plus, il est possible d'éditer n'importe quel type de dictionnaire s'il est encodé en XML.

Plusieurs projets de construction de ressources lexicales ont utilisé ou utilisent toujours cette plate-forme avec succès. C'est le cas par exemple du projet GDEF de dictionnaire bilingue estonien-français⁷. Le code de cette plate-forme est disponible gratuitement en source ouverte en téléchargement depuis la forge du laboratoire LIG⁸.

4 Description de la ressource à construire

4.1 Microstructure des articles basée sur la théorie Sens-Texte

La microstructure des articles composant les volumes monolingues est une simplification de celle du projet Papillon. Chaque article est cette fois basé sur le vocable. Un vocable étant soit un regroupement de lexies (sens de mot), soit une locution.

Pour faire face aux niveaux de compétence différents selon les contributeurs, l'interface d'édition pourra s'adapter et afficher une granularité d'information adaptée. Par exemple, un contributeur débutant sera invité à renseigner une simple glose pour caractériser une lexie, alors qu'un linguiste expert devra décrire une formule sémantique complète. De même, certains contributeurs seulement auront accès à la liste des fonctions lexicales à remplir. La figure 4 décrit l'article correspondant au vocable « abandonner ». Celui-ci n'a pour l'instant qu'un seul sens, caractérisé par une glose (*laisser tomber un truc*) et une formule sémantique (action sur un objet : humain ou animal X ~ entité Y).

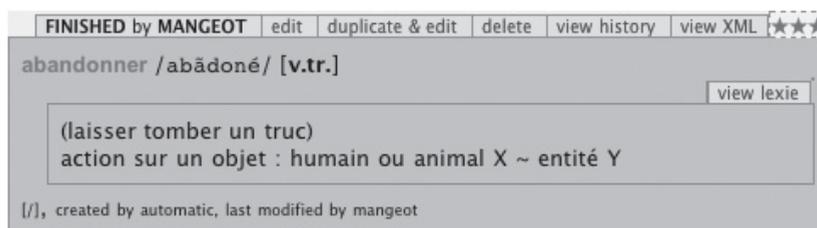


Figure 4

4.2 Macrostructure pivot via des interfaces bilingues

La macrostructure est également tirée du projet Papillon avec un volume monolingue pour chaque langue et un volume pivot au centre. Cependant, afin de ne pas dérouter les utilisateurs, ceux-ci contribueront via une interface présentant une vue classique de dictionnaire bilingue. Chaque lien bilingue langue A → langue B

⁷ www.estfra.ee

⁸ jibiki.ligforge.imag.fr

ajouté via cette interface sera en fait traduit en arrière-plan par la création de deux liens interlingues ainsi que d'une axie représentant le lien de traduction d'origine pour obtenir finalement : langue A → axie pivot → langue B (voir figure 5).

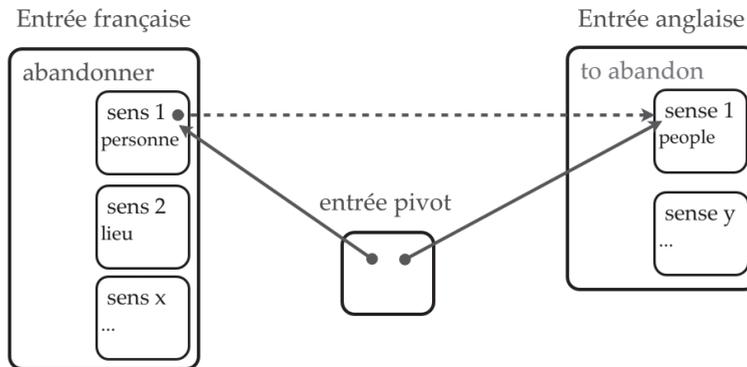


Figure 5

4.3 Établissement des liens bilingues et interlingues

Lorsqu'un contributeur veut ajouter un lien de traduction entre un vocable V_a de langue A et un vocable V_b de langue B, il peut établir ce lien à différents niveaux.

La solution idéale est de relier un sens de mot S_a du vocable V_a à un autre sens de mot S_b du vocable V_b . Dans ce cas, le lien est bijectif et S_b est donc aussi relié à S_a .

Si le vocable V_b n'a pas encore de sens de mots précis ou si le contributeur n'arrive pas à choisir de sens de mot, il peut relier S_a directement au vocable V_b . Dans ce cas, un nouveau sens de mot S_b' est créé avec un niveau de qualité brouillon et le lien ainsi que les sens de mots sont marqués comme étant à raffiner.

Dans le cas de la récupération de données existantes, il est bien souvent impossible de rattacher une information à un sens de mot précis. Dans ce cas, on ajoute à la fin du vocable V_a l'information selon laquelle un des sens de mots de V_a peut être relié à un sens de mot de V_b , mais cette information ne sera pas ajoutée à V_b . Elle sera bien sûr marquée comme étant à raffiner d'urgence!

Grâce à la macrostructure pivot, si deux liens langue A → langue B et langue B → langue C existent, alors il sera automatiquement créé un lien langue A → langue C dont le niveau sera de qualité brouillon et marqué comme à réviser (voir figure 6).

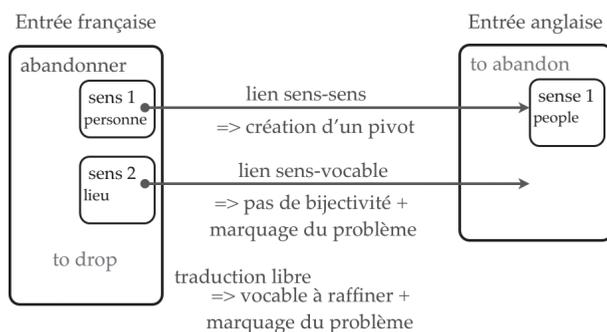


Figure 6

4.4 Niveaux de qualité des données et des contributeurs

Chaque partie d'information de chaque article se verra attribuer un niveau de qualité. Les niveaux s'échelonnent de 1 étoile pour un brouillon (données récupérées dont la qualité n'est pas connue) à 5 étoiles, qualité certifiée par un expert (par exemple, un lien de traduction validé par un traducteur assermenté). La figure 4 montre par exemple que l'article du vocable « abandonner » est de niveau 3 étoiles.

De la même manière, les contributeurs se verront assigner un niveau de compétence (1 à 5 étoiles également). 1 étoile étant le niveau d'un débutant inconnu dans la communauté et 5 étoiles étant le niveau d'un expert reconnu.

Ensuite, lorsqu'un contributeur de niveau 3 révisé un article de niveau 2, l'article monte automatiquement au niveau 3. De même, si le travail d'un contributeur est systématiquement validé sans corrections par d'autres contributeurs de niveau supérieur, celui-ci peut passer automatiquement au niveau supérieur au bout d'un certain seuil (par exemple 10 contributions).

Pour aller plus loin, nous envisageons d'analyser le travail des contributeurs. Si une personne contribue massivement par exemple sur un domaine particulier, le système pourra de manière automatique lui envoyer régulièrement des propositions de contribution dans son domaine.

5 Méthodologie d'élaboration des données

La méthodologie d'élaboration des données est constituée de trois étapes principales : la récupération de données existantes, la collecte de données via des jeux sérieux et enfin la contribution en ligne sur le Web.

5.1 Récupération des données existantes

Afin d'encourager les contributions, il est préférable de proposer un squelette de dictionnaire à modifier plutôt qu'un dictionnaire vide (syndrome de la page blanche). Pour chaque langue en présence, une liste de mots de cette langue sera récupérée

pour créer une première liste d'articles. Il sera toujours possible de créer un nouvel article, mais les créations seront soumises à vérification.

Selon les sous-projets et les langues en présence, plusieurs dictionnaires peuvent être utilisés :

- les dictionnaires des projets Fe* (français-anglais + autre langue): FeM (malais) (Gut et al. 1996), FeT (thaï), FeV (vietnamien) ;
- la base DiCo pour le français ;
- le dictionnaire bilingue français-vietnamien VietDict.

La figure 7 montre un exemple d'article du dictionnaire Fe* avec des traductions en anglais, vietnamien (gris clair), malais (noir) et thaï.

| FINISHED by edit duplicate & edit delete view history view XML | | | | | |
|--|--------------------------------------|---|---------------------|------------------|---|
| abandonner | /ABAN-DONE-/ | | | | |
| [v.tr.] | | | | | |
| (laisser) | to leave | bỏ, không nhìn nhận | meninggalkan | ทิ้ง | ทอดทิ้ง |
| <i>cette mère a abandonné ses enfants</i> | <i>this mother left her children</i> | | | | <i>người mẹ này đã bỏ rơi những đứa con của mình.</i> |
| | | ibu ini telah meninggalkan anak-anaknya | | | แม่คนนี้ได้ทอดทิ้งลูกๆของเธอ |
| (désert(er) (milit.) | to desert | bỏ đi, trốn đi, dạo vòng | meninggalkan tugas | | หนีทหาร |
| (renoncer à) | to give up | to abandon | từ bỏ, khước từ | melepaskan | membiarkan |
| | | | | | ยกเลิก ล้มเลิก ละทิ้ง |
| <i>il a abandonné son projet</i> | <i>he had gave up his project</i> | <i>cậu ta đã bỏ dự định của mình.</i> | | | <i>dia telah membiarkan proyeknya</i> |
| | | เขาได้ล้มเลิกโครงการของเขา | | | |
| (céder) | to give in to | để lại, bán đi, nhượng lại | mengalah | menyerah | ยอม |
| s'abandonner | /SABAN-DONE-/ | | | | |
| [v.pr.] | | | | | |
| <i>s'abandonner à</i> | <i>to give oneself up to</i> | <i>đi xuống, suy sụp, sụp đổ</i> | <i>menyerah kpd</i> | <i>ยอมให้กับ</i> | <i>ยอมรับ</i> |
| <i>elle s'est abandonnée au désespoir</i> | <i>she has given up to despair</i> | <i>cô ta đã rơi vào tuyệt vọng</i> | | | <i>dia telah menyerah kpd kekecewaan</i> |
| | | เธอยอมรับความผิดหวัง | | | |

Figure 7

Le nombre d'étoiles initial des articles générés à partir de ces données est fixé en fonction de la qualité du dictionnaire et de la granularité des données récupérées.

Pour la langue khmère, il existe à l'heure actuelle un dictionnaire français-khmer informatisé (Richer *et al.* 2007) qui, commencé à la fin des années 90, a été achevé en 2006 par un petit groupe de chercheurs informaticiens réunis dans l'association « Pays Perdu » créée par Denis Richer, ethnolinguiste français établi à Siem Reap (Cambodge). Cette première version du dictionnaire a été publiée au printemps 2007 et comporte 20 000 articles. Le dictionnaire est au format Word. La partie khmère consiste en une simple transcription phonétique khmère de la traduction du mot-vedette écrite dans une police spéciale API (SILSophia IPA93) créée par le *Summer Institute of Linguistics*. La figure 8 montre à gauche un extrait de ce dictionnaire et à droite ce que l'on voudrait obtenir (écriture khmère au lieu de la transcription phonétique).

| French | Khmer | | |
|------------|-----------------|--------|---------------|
| jarret | kōnlēak-cōl | jarret | កន្លាក់-រើង |
| jars | keḷān-chmōl | jars | កាន-ឈ្មាស |
| jasmin | rnlih | jasmin | ផ្លែ |
| jauge | ḷcamnoh-rōḷvuel | jauge | កំនត់-រង្វាល់ |
| — (techn.) | māet-stueḷ | | |

Figure 8

5.2 Collecte de données via les jeux sérieux

Il s'agira ici de lancer un JeuxDeMots pour chaque langue du projet. Le JeuxDeMots français est déjà lancé depuis maintenant 2 ans. Le JeuxDeMots vietnamien est déjà traduit et vient d'être lancé officiellement⁹. Le JeuxDeMots khmer est en cours de traduction. Nous espérons trouver un succès comparable au JeuxDeMots français. Pour aller plus loin, il faudrait réfléchir à des jeux permettant la collecte de données bilingues.

5.3 Contribution en ligne sur le Web

Les données récupérées et collectées sont ensuite fusionnées pour donner naissance à un squelette de dictionnaire. Celui-ci est ensuite mis en ligne pour correction et enrichissement.

Conclusion

Le projet est déjà relativement bien avancé. La plupart des aspects techniques concernant la plate-forme et les jeux sérieux en ligne sont réglés. Il reste maintenant à récupérer et convertir les ressources existantes. L'enjeu majeur du projet se trouve en fait dans notre capacité à motiver les communautés de contributeurs. Nous espérons que notre expérience et l'attrait d'un tel projet nous permettront d'avancer sur ces aspects sociologiques.

Les retombées d'un tel projet seront nombreuses et contribueront à relancer l'intérêt de la francophonie dans les pays de l'Asie du Sud-Est. Les données produites pourront être utilisées par les apprenants du français dans ces pays, ou par les francophones souhaitant s'initier à une langue d'Asie du Sud-Est. Les dictionnaires pourront être consultés directement en ligne ou sur PDA par les touristes ou les hommes d'affaires.

Les communautés de contributeurs permettront de lancer une dynamique de coopération autour d'un but humaniste, voire humanitaire. De plus, cela peut susciter l'intérêt pour l'apprentissage des langues traitées dans le projet et l'élargissement à d'autres langues de la région.

⁹ choichu.mica.edu.vn

Remerciements

Le projet MotÀMot est financé en partie par une Action de Recherche en Réseau du Réseau Lexicologie, Terminologie, Traduction (ARR-LTT) de l'Agence Universitaire de la Francophonie (AUF).

Bibliographie

- Berment (V.), 2004 : *Méthodes pour informatiser des langues et des groupes de langues peu dotées*. Thèse de nouveau doctorat, Université Joseph Fourier, Grenoble, France.
- Blanc (E.), 1996 : « Une maquette de base lexicale multilingue à pivot lexical : parax. *Lexicomatique et Dictionnairique* », dans *Actes du colloque LTT, Lyon, septembre 1995*, aupelf-uref, Montréal, Canada, p.43-58.
- Caelen-Haumont (G.), 2009 : *Prosodie et sens: une approche expérimentale*, L'Harmattan-Marges Linguistiques.
- Gut (Y.), Megat Ramli (P.R.), Zaharin (Y.), Chuah Choy (K.), Samat (S.A.), Boitet (C.), Nédobejkine (N.), Mathieu Lafourcade (M.) *et al.*, 1996 : *Kamus Perancis-Melayu Dewan, dictionnaire français-malais*. Dewan Bahasa Dan Pustaka, Kuala Lumpur, 667 p.
- Lafourcade (M.) et Joubert (A.), 2008 : « JeuxDeMots: un prototype ludique pour l'émergence de relations entre termes », dans *JADT 2008 : 9^{es} Journées internationales d'Analyse statistique des Données Textuelles*, Lyon, France, p.657-666.
- Mangeot (M.), 2001 : *Environnements centralisés et distribués pour lexicographes et lexicologues en contexte multilingue*, thèse de nouveau doctorat, spécialité informatique, Université Joseph Fourier (Grenoble 1), septembre 2001, 280 p.
- Mangeot (M.), Sérasset (G.) et Lafourcade (M.), 2003 : « Construction collaborative de données lexicales multilingues, le projet Papillon », dans Zock (M.) et Carroll (J.), éd., *Revue tal, édition spéciale, Les dictionnaires électroniques: pour les personnes, les machines ou pour les deux? (Electronic dictionaries: for humans, machines or both?)*, Vol. 44, p.151-176.
- Mel'čuk (I.) et Polguère (A.), 2006 : « Dérivations sémantiques et collocations dans le DiCo/laf », dans Blumenthal (P.) et Hausmann (F. J.), dir., *Langue française, numéro spécial sur la collocation « Collocations, corpus, dictionnaires »*, n° 150, juin 2006, p.66-83.
- Polguère (A.), 2006 : « Structural properties of Lexical Systems: Monolingual and Multilingual Perspectives », dans *Proceedings of the Workshop on Multilingual Language Resources and Interoperability (coling/acl 2006)*, Sydney, p.50-59.
- Richer (D.), T. Keo (T.), Vania (I.), 2007 : *Dictionnaire français-khmer (en phonétique)*, D.R. Édition, ISBN-13: 890-0-9.
- Sam (S.), 2006 : *Analyse de la langue khmère en vue de la synthèse de la parole*, M2R (DEA) de l'Université Joseph Fourier, Grenoble, France.
- Sam (S.), Éric Castelli (E.), L. Protin (L.), Pham Thi (N.Y.), 2007 : *Traitement automatique de la langue khmère: rapport scientifique final* — Projet AUF TALK ITC — MICA — Cambodge, mai 2007.
- Seng (S.), 2007 : *Collecte et traitement de données en vue de la reconnaissance automatique de la parole en langue khmère*, Rjcp 2007, Paris, France, à paraître.

Sérasset (S.), Brunet-Manquat (F.) *et al.*, 2006 : « Multilingual legal terminology on the Jibiki platform: the LexALP project », dans *Coling-ACL 2006 : Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, Sydney.

Sérasset (G.), 1994 : *SUBLIM: un Système Universel de Bases Lexicales Multilingues, et NADIA, sa spécialisation aux bases lexicales interlingues par acceptions*, Thèse de nouveau doctorat, spécialité informatique, Université Joseph Fourier (Grenoble 1), 194 p.

Valeria Zotti,
Dipartimento di Lingue e Letterature Straniere Moderne,
Università di Bologna,
Bologna,
Italie,
valeria.zotti@unibo.it.

Pour une réinterprétation du dictionnaire bilingue face à la pluralité linguistique de l'espace francophone : l'exemple du français québécois

Résumé

L'article expose les stratégies mises en œuvre au niveau des informations macro- et microstructurelles dans un nouveau dictionnaire bilingue qui s'ouvre à la pluralité linguistique de l'espace francophone. On y présente les innovations introduites pour améliorer le traitement de l'un des exemples emblématiques d'écart par rapport à la norme parisienne : le français québécois.

Mots-clés : lexicologie ; lexicographie ; traduction ; variation géographique ; Québec.

Introduction

En s'ouvrant à la pluralité linguistique de la francophonie, le dictionnaire bilingue a récemment acquis un rôle et une fonction inédite : « non plus réservoir d'une unité utopique de la langue française » (Zotti 2007 : 97), mais lieu de rencontre de ses multiples visages.

Le français québécois est l'un des exemples les plus emblématiques de cette ouverture à la diversité linguistique. Implanté sur le territoire canadien dès le XVII^e siècle, il est devenu au Québec une entité fortement identitaire, autonome par rapport au français de France. Du point de vue numérique, le français québécois est prééminent : pour l'ensemble du Québec, les statistiques linguistiques 2001 ont évalué à 5 802 020 le nombre de « vrais » francophones, ayant le français comme langue maternelle (StatCan 2009). En plus, depuis longtemps, Montréal est la seconde ville francophone du monde, après Paris et à égalité avec Bruxelles (Walter 1988 : 220).

La mondialisation au sens large, l'essor des nouvelles technologies, en particulier de la Toile, et le développement des transports facilitant les déplacements hors de l'UE font en sorte qu'aujourd'hui enseigner et pratiquer la langue française ne peut plus se limiter à la seule « variété » de France. L'accès à la compréhension du français québécois s'avère nécessaire aussi pour une population non francophone. Le

dictionnaire bilingue, dont la fonction première est la mise en contact de deux différents systèmes linguistiques, se doit d'assumer convenablement ce nouvel enjeu.

Cependant, jusqu'à la fin du xx^e siècle, l'intégration du français québécois dans les nomenclatures des dictionnaires bilingues s'est limitée à l'introduction d'un échantillon représentatif de québécismes formels. Aujourd'hui, l'acceptation de la légitimité désormais acquise par le parler franco-canadien devrait conduire à la description aussi de ses particularités lexicales, sémantiques et morphosyntaxiques, traditionnellement considérées à mauvais escient comme des déviations par rapport à la norme.

Après avoir analysé la place du français québécois dans les principaux dictionnaires bilingues français-italien/italien-français contemporains (Zotti 2007), dans cette communication nous présenterons la suite de notre recherche, à savoir l'expérience menée pour aménager le traitement des mots de la francophonie dans un nouveau dictionnaire bilingue (NDB) en cours de réalisation.

Quatre interventions principales ont été envisagées dans ce nouveau dictionnaire : 1. l'établissement de critères plus cohérents pour la sélection des québécismes dans la nomenclature ; 2. la correction des traduisants italiens proposés à travers la consultation de ressources lexicologiques québécoises ; 3. l'introduction d'une collection d'exemples qui illustrent à la fois la société et la culture québécoises ainsi que des usages caractéristiques qui relèvent de la morphosyntaxe ; 4. la structuration XML des entrées afin de réaliser en même temps deux versions, papier et électronique, chacune adaptée à un public cible différent.

1 Des critères plus cohérents pour la sélection de la nomenclature

1.1 Les sources actuelles de la nomenclature

Comme nous l'avons déjà remarqué (Zotti 2003 : 389) : « Traditionnellement chaque dictionnaire bilingue s'apparente à un ou deux dictionnaires monolingues ». Les nomenclatures des dictionnaires de langue servent de sources pour la constitution des nomenclatures des dictionnaires bilingues. Dès lors, la sélection des québécismes à intégrer dans les dictionnaires bilingues est, elle aussi, faite sur des bases qui n'assurent pas la cohérence interne du dictionnaire.

Après avoir dépouillé les nomenclatures des éditions 2000 et 2007 de trois dictionnaires bilingues contemporains (Boch, Paravia, Garzanti) en relevant le nombre de québécismes présents (voir Image 1 : Graphique), on remarque que l'augmentation des entrées portant sur des emplois québécois dans chaque dictionnaire bilingue coïncide avec la hausse enregistrée à l'intérieur de chaque dictionnaire de langue correspondant.

Bien que le dictionnaire de langue *Le Petit Robert* ait été le pionnier quant à la quantité de formes et de sens québécois introduits dans l'édition 1967 (Dugas 1979),

aujourd'hui la primauté quantitative appartient au *Petit Larousse* avec ses près de 470 dans l'édition de 2005 (Poirier 2003 : 201-202).

Le *Garzanti*, qui hérite sa nomenclature du dictionnaire *Le Petit Larousse*, intègre en effet le plus grand nombre de québécismes. Sa dernière édition (2007) enregistre une augmentation considérable qui s'insère parfaitement dans le sillage de la fournée de mots francophones de 2005. Cependant, dans la nomenclature du *Garzanti* 2007 on repère 289 québécismes, alors que beaucoup plus sont présents dans l'édition 2005 du *Petit Larousse*. Cette donnée indique qu'une sélection d'entrées est faite pour établir les nomenclatures des dictionnaires bilingues.

Quels critères président à cette sélection ? D'une analyse ponctuelle des nomenclatures, il ressort que dans les trois bilingues « une considérable diversification, quant à la sélection et la répartition de la nomenclature, est observable » (Zotti 2007 : 51-52). Cependant, à partir des contrôles effectués dans plusieurs éditions du dictionnaire *Le Petit Robert*, il ressort que plusieurs québécismes enregistrés dans Boch 2007 y font défaut. L'impression d'ensemble est que la sélection a été faite de façon arbitraire, sans que l'on ait tenu compte attentivement des spécificités du dictionnaire bilingue.

1.2 Des sources ciblées pour un nouveau dictionnaire bilingue

Le dictionnaire bilingue a une fonction (décodage/encodage) et une direction (public cible français/autre langue) différente par rapport au dictionnaire de langue. Dans cet outil, la « norme », bien que représentée, est mise légèrement de côté au profit de la description de l'« usage », qu'il soit contemporain ou non. Ce que le lexicographe bilingue doit prendre en considération pour la sélection de la nomenclature, c'est l'utilisation effective de la langue dans les deux systèmes linguistiques mis en parallèle, à savoir la probabilité que l'utilisateur du dictionnaire bilingue rencontre des occurrences des mots enregistrés.

La taille du dictionnaire se construit donc sur la base d'une variable fondamentale : le public cible. Jusqu'à présent, dans le panorama franco-italien, le public cible d'un dictionnaire général bilingue, c'est-à-dire d'un outil pouvant contenir entre 200 000 mots et locutions (Paravia 2007) et 350 000 mots et acceptions environ (Boch 2007 et Garzanti 2007), comprend un public très vaste : « enseignants, lycéens et étudiants, traducteurs et interprètes, spécialistes de toutes les disciplines » (Boch 2007 : préface).

Dans d'autres paysages lexicographiques, comme l'Angleterre, l'offre est plus diversifiée. À côté du « dictionnaire général bilingue », il existe un autre type de dictionnaire qu'on appelle généralement « grand dictionnaire », souvent en deux volumes (il peut contenir environ 400 000 mots et expressions), qui s'adresse surtout aux professionnels (« traducteurs et enseignants ») l'utilisant davantage pour la production-encodage en L2.

Dans le nouveau dictionnaire bilingue sur lequel nous travaillons, les deux catégories à peine cernées s'appliquent en même temps aux deux supports envisagés, le support papier et le support électronique. Cette double nature permet d'adapter et de personnaliser la taille du dictionnaire en fonction du public cible de référence.

Pour d'évidentes raisons d'espace, le support papier requiert une sélection plus poussée de la nomenclature. Seule la langue générale, à savoir les mots les plus fréquemment employés aujourd'hui dans la vie courante, est représentée selon une perspective synchronique, le public cible principal étant l'apprenant. On veille pour cela à ce que des mots peu employés (vieux) ou circonscrits à certains domaines (techniques) sortent de la nomenclature générale.

Le support électronique, offrant de vastes possibilités d'inclusion, correspond à la taille d'un « grand dictionnaire ». Tout ce qu'on trouve dans la version papier y est inclus, mais on pourra se permettre de ratisser large sur les mots francophones, en les décrivant dans différentes situations énonciatives : presse écrite générale et spécialisée, textes techniques, conversations informelles, etc. La perspective adoptée est synchronique, diachronique et terminologique, le public cible étant principalement le spécialiste de la L2 (l'enseignant ou le traducteur professionnel littéraire ou technique).

Dans les deux cas, l'exploration du web comme corpus de l'« usage » sert à vérifier la fréquence des mots. Le web constitue une source infinie d'informations et de données linguistiques représentatives des langues contemporaines et caractérisées par la présence d'une grande variété de textes du point de vue thématique et stylistique (Emirhanian et Fouqueré 2003 : 8). Grâce à ses qualités intrinsèques et toutes précautions prises, à savoir la vérification des données dans des ressources lexicologiques québécoises (voir points 2 et 3) et la collaboration d'informateurs francophones natifs du Québec, le web constitue un lieu de contrôle essentiel de l'actualité de la nomenclature à intégrer dans la version papier et électronique de notre nouveau dictionnaire bilingue.

2 La correction des traduisants

Dans les dictionnaires bilingues contemporains, on relève que la méconnaissance des écarts sémantiques du français du Québec par rapport au français de France engendre souvent des traductions erronées. On donnera d'abord quelques exemples représentatifs de ces fautes de traduction et, ensuite, on illustrera comment elles peuvent être corrigées à travers la consultation des ressources offertes par l'équipe du *Trésor de la langue française au Québec* : notamment, l'*Index lexicologique québécois* (ILQ 2009) et la *Base de données lexicographiques panfrancophone-Québec* (BDLP 2005).

L'entrée *blonde* constitue un exemple particulièrement représentatif de fautes ou imprécisions fréquemment repérables dans les dictionnaires bilingues contemporains.

[1] (Paravia 2007):

blonde *f.* [...] can. (*petite amie*) *biondina*

(Garzanti 2007):

blonde [blōd] n.f. [...] (*Canada*) (*fam.*) *ragazza molto corteggiata*

Le *Paravia* (2007) propose l'équivalent italien *biondina*, qui signifie en italien simplement « jeune fille blonde ». À la place d'un traduisant lexical, le *Garzanti* (2007) propose une autre équivalence: la glose définitionnelle *ragazza molto corteggiata*, qui signifie « fille très courtisée ». Les deux équivalents traductionnels proposés ne sont pas satisfaisants. Il paraît évident que le traitement de l'entrée *blonde* dans les deux dictionnaires bilingues analysés a été fait à distance, sans que les lexicographes, sans doute italiens, aient contrôlé avec soin le sens du mot au Québec. La consultation des ressources lexicologiques et lexicographiques québécoises s'avère essentielle pour résoudre ces problèmes de compréhension du sens et de traduction.

2.1 L'Index lexicologique québécois

La consultation de l'*ILQ* permet de trouver dans quelles ressources linguistiques québécoises (glossaires, manuels de bon usage, listes de mots, dictionnaires, articles dans des revues, chroniques de langage, thèses, mémoires, etc.) un mot ou une expression a été l'objet d'étude ou d'un commentaire depuis le milieu du XVIII^e siècle jusqu'à nos jours.

Pour le mot *blonde*, l'*ILQ* donne 31 relevés dans 15 références bibliographiques différentes (voir Image 2: « blonde »). Parmi les différentes sources indiquées, nous avons consulté en particulier les sources n. 3 (Bergeron 1981), n. 6 (Clapin 1894) et n. 15 (Poirier P. 1993).

[2] (Bergeron 1981):

Blonde *n.f.* – Jeune fille qui a gagné l'affection d'un jeune homme. – Jeune fille courtisée. Ex.: *Ce soir je vais voir ma blonde.*

(Clapin 1894):

Blonde *s.f.*, Fiancée, promise, amante, maitresse, etc.: – C'est sa *blonde*, c.-à-d. c'est sa promise, sa bonne amie [...]. Au Canada, par contre, on entend par *blonde* toute jeune fille quelconque à laquelle on fait la cour.

(Poirier P. 1993):

Blonde. *Subst. fém.* Amante, promise, bonne amie. Se dit d'une brune aussi bien que d'une *blonde*; mais la *blonde*, au teint clair est en Acadie comme chez les paysans de France, le type préféré de beauté féminine: *aller voir sa blonde, veiller avec sa blonde.* [...]

La consultation de ces ressources lexicographiques monolingues nous permet de découvrir qu'en français québécois le mot *blonde* a plusieurs acceptions qui diffèrent de la signification du mot en France. Le traduisant italien proposé par *Paravia* (2007), *biondina*, se révèle erroné, car il équivaut au sens français de France. La traduction proposée par *Garzanti* (2007), *ragazza molto corteggiata*, se révèle par contre correcte, mais insuffisante. Elle ne se rapporte qu'aux définitions « Jeune fille courtisée » (Bergeron 1981) et « toute jeune fille quelconque à laquelle on fait la cour » (Clapin 1984). Les trois sources consultées enregistrent une autre acception, « fiancée, promise », placée en première place dans l'article, donc vraisemblablement la plus courante. La consultation d'un dernier ouvrage lexicographique de référence (*DHFQ* 1998), qui étonnamment ne figure pas parmi les sources relevées par *l'ILQ*, permet d'effacer ce doute.

[3] (*DHFQ* 1998):

Blond, Blonde *adj. et n.* [...] 2° *N.f.* Jeune fille courtisée, recherchée en mariage; amoureuse, fiancée. *Par ext.* Amante, amie, passagère, compagne. [...] Général au Canada français au début du siècle (FSPFC), encore usuel partout de nos jours. 2 Blonde évoque le plus souvent l'idée d'une jeune fille avec laquelle on entretient une relation amoureuse stable, conduisant normalement au mariage; il s'agit de l'emploi traditionnel du mot au Canada français (qui fait pendant à celui de cavalier, en parlant de l'homme). Le mot s'applique également, déjà au XIX^e s., à l'amie passagère, à la femme avec laquelle on ne veut pas s'engager sérieusement. De nos jours, *blonde* prend à l'occasion une connotation amusante (il est en grande faveur auprès des enfants) et son emploi n'est guère possible lorsqu'on fait des présentations. Par ailleurs, le mot tend à s'employer en parlant de la femme avec laquelle on vit (parallèlement à *chum*, en parlant de l'homme.) [...]

Toutes les attestations fournies par le *DHFQ* (1998) confirment que la première acception du mot était employée surtout au début du siècle dernier, alors que la deuxième est encore usuelle de nos jours. La première acception mérite donc d'être décrite dans un dictionnaire bilingue.

Nous proposons ci-après le premier stade de rédaction de l'entrée *blonde* de notre nouveau dictionnaire bilingue. On y signale les traduisants corrects disposés selon le critère de fréquence d'usage, du plus au moins fréquent. La marque d'usage en italien entre crochets signale que la deuxième acception est vieillie.

[4] (NDB):

blonde *n.f.* queb.

1 (*petite amie, compagne*) ragazza, compagna

2 (*fiancée, promise*) [desueto] promessa (sposa)

2.2 La Base de données lexicographiques panfrancophone-Québec

La consultation d'une autre ressource, la *BDLP-Québec* (2005), est très fructueuse pour recueillir un grand nombre d'informations lexicographiques sur le fonctionnement linguistique de mots mal décrits dans les dictionnaires bilingues analysés. Il s'agit d'un fonds informatisé résultant d'une analyse de corpus qui réunit les données issues des travaux de l'équipe du *Trésor de la langue française au Québec* (Université Laval, Ciral) sur les mots et les emplois de la langue générale au Québec. Pour illustrer l'utilité de la consultation de cette base, nous avons choisi le québécisme sémantique *achalandage*, qui apparaît dans un seul dictionnaire bilingue contemporain (Boch 2007).

[5] (Boch 2007):

achalandage s.m. (quebec.) affluenza (f.): un achalandage de mille personnes par jour, un'affluenza di mille persone al giorno.

La *BDLP-Québec* donne accès à un grand nombre d'informations: la définition de la vedette, qui illustre sa signification au Québec; des exemples courts présentés sous forme de syntagmes; des citations tirées de corpus littéraires, journalistiques ou scientifiques; des renvois aux dérivés, composés ou mots de la même famille que le mot entrée; la définition de la vedette en français de référence; pour finir, l'information étymologique (voir Image 3 « *achalandage* »).

L'information différentielle sur la définition du mot au Québec et en France est la plus précieuse pour le lexicographe bilingue car elle lui permet de savoir que l'entrée en question enregistre une variation sémantique sur le plan diatopique. À partir de cette information contrastive, il est ainsi plus aisé d'identifier le traduisant italien le plus approprié pour rendre compte des écarts sémantiques intralingaux.

Les informations recueillies nous permettent de vérifier que, dans ce cas, le traduisant *affluenza* proposé par (Boch 2007) pour le mot *achalandage* est correct. Nous avons ensuite précisé l'emploi de ce mot au Québec par le biais d'un exemple qui fournit aussi à l'utilisateur italien des informations extralinguistiques sur la culture et la société canadiennes. L'exemple donné dans l'article de notre nouveau dictionnaire bilingue, tiré du corpus de la *BLDP-Québec* et librement adapté, mentionne deux zones géographiques du Canada: « l'Ontario central » et « la vallée du Saint-Laurent ».

[6] (NDB):

achalandage n.m. queb. affluenza (f.) L'achalandage de voitures entre les villes de l'Ontario central est moins intense que dans la vallée du Saint-Laurent, L'affluenza di auto tra le città dell'Ontario centrale è meno intensa rispetto alla valle del San Lorenzo.

3 L'aménagement morphosyntaxique

Aujourd'hui, il est possible d'admettre qu'« il existe au Québec une norme du français écrit et parlé propre à satisfaire les besoins de communication nationaux, internationaux et francophones » (Boulangier 1998 : 130). Dans notre nouveau dictionnaire, nous avons choisi d'attester la norme de l'usage courant québécois, en introduisant aussi ses particularités morphosyntaxiques. Cela est impératif dans un dictionnaire bilingue qui suppose des usagers plus en contact avec les pays étrangers, et qui, par conséquent, peuvent se trouver plus facilement dans la condition de confondre cet usage avec un emploi erroné. Ainsi, un énoncé du type « J'ai *passé à travers* l'épreuve malgré un bien *achalant* rhume » est partiellement opaque pour un locuteur étranger ou non québécois, à cause de la difficulté lexicale posée par le québécisme formel « achalant », mais surtout à cause de l'emploi apparemment fautif de la préposition « à » dans l'expression « passer à travers ».

Pour illustrer la norme de l'usage québécois, la consultation des ressources lexicologiques du *Trésor de la langue française au Québec* se révèle encore une fois précieuse. Par la recherche assistée de l'ILQ, il est possible d'accéder, non seulement aux mots, mais aussi aux formes ou expressions, et de vérifier si elles sont acceptables au Québec (voir Image 4 : « passer à travers »).

En nous basant sur les remarques de Boysen (1990 : 47-48), nous avons déjà affirmé qu'« il serait irréaliste de prétendre à l'inclusion de toutes les particularités morphosyntaxiques du français québécois » (Zotti 2007 : 82), étant donné que la multiplicité de formes que, par exemple, la seule préposition « à » prend au Québec rendrait très complexe son intégration à l'intérieur d'un article bilingue. Pour illustrer ces écarts morphosyntaxiques, nous avons alors opté pour l'introduction d'exemples reflétant l'usage québécois. En particulier, les citations tirées du corpus qui sont données par la *BDLP-Québec* permettent d'insérer le mot vedette dans une dimension discursive. Ces formes et expressions ne seront pas repérables dans la nomenclature, mais la consultation plein texte de la version électronique permettra de les retrouver dans le corps des articles.

Ainsi, dans l'entrée *achalant* de notre dictionnaire, nous proposons comme exemple l'énoncé en partie opaque donné plus haut, qui illustre en discours à la fois le sens du mot entrée et l'expression *passer à travers* :

[7] (NDB):

achalant, ante *agg.* queb. (*fam.*) fastidioso *J'ai passé à travers l'épreuve malgré un bien achalant rhume*, Ho superato la prova nonostante un fastidiosissimo raffreddore.

Dans l'article *blonde* aussi, l'ajout d'exemples tirés de corpus ou de citations nous permet de montrer des emplois morphosyntaxiques québécois. Le niveau phrasique et idiomatique est mis en jeu à travers le choix d'illustrer dans l'exemple l'emploi de

la forme *tomber en amour*, employée couramment au Québec et empruntée à l'anglais *fall in love*. L'aspect différentiel est mis en relief typographiquement en soulignant les structures qui différencient le français du Québec du français de France.

[8] (NDB):

blonde *n.f.* queb.

1 (*petite amie, compagne*) ragazza, compagna *Ma blonde et moi on est tombé en amour au moment où on s'est vu*, La mia ragazza ed io ci siamo innamorati a prima vista

2 (*fiancée, promise*) [*desueto*] promessa (sposa) *Quand vous partez pour ces pays étrangers, Si vous avez quelque blonde à conserver, Allez les voir auparavant que de partir* (*Chansons de voyageurs*, 1856)

4 La structuration XML des entrées

La conception du nouveau dictionnaire bilingue que nous réalisons s'inscrit dans un programme rédactionnel préétabli qui a une incidence sur la structuration formelle des entrées. Afin de satisfaire les besoins des deux types de public cible envisagés (voir point 1.2), nous avons envisagé de réaliser en même temps la version papier et électronique. Chaque article est rédigé sur une microstructure élargie qui, seulement dans un second moment, sera modelée et personnalisée pour la réalisation de chaque version.

En raison de sa malléabilité vers tout genre lexicographique, nous avons adopté le modèle lexicographique standard X_mLex (Le Meur et Derouin: 2006). Nous présentons ci-après une version extrêmement simplifiée de la structuration au format X_mLex de l'article *blonde*, divisé en trois blocs principaux (HeadwordBlock, 1SenseBlock, 2SenseBlock). Dans la version papier de l'article, sera affichées seulement la rubrique contenant l'entrée et les informations la concernant (HeadwordBlock) et la première rubrique sémantique (1SenseBlock). Un script permettra de ne pas afficher les qualificatifs sémantiques lorsqu'un seul bloc sémantique apparaît. L'article de la version électronique sera par contre composé de l'ensemble des informations microstructurelles (voir entrée *blonde* donnée plus haut, point 3).

[9] Version X_mLex

<DictionaryEntry>

<HeadwordBlock>

<Headword> blonde </Headword>

<PartOfSpeech> n </PartOfSpeech>

<GrammaticalGender> f </GrammaticalGender>

<GeographicalUsage> queb </GeographicalUsage>

</HeadwordBlock>

```

<SenseBlock>
  <SenseQualifier> petite amie </SenseQualifier>
  <SenseQualifier> compagne </SenseQualifier>
  <TranslationBlock>
    <Translation> ragazza </Translation>
    <Translation> compagna </Translation>
  </TranslationBlock>
</Example>
  <SourceLanguage> Ma blonde et moi on est tombé en amour au
  moment où on s'est vu </SourceLanguage>
  <TargetLanguage> La mia ragazza ed io ci siamo innamorati a prima
  vista </TargetLanguage>
</Example>
</SenseBlock>
<SenseBlock>
  <SenseQualifier> fiancée </SenseQualifier>
  <SenseQualifier> promise </SenseQualifier>
  <Frequency> desueto </Frequency>
  <Translation> promessa (sposa) </Translation>
  <Citation> Quand vous partez pour ces pays étrangers, Si vous avez quel-
  que blonde à conserver, Allez les voir auparavant que de partir (Chansons
  de voyageurs, 1856) </Citation>
</SenseBlock>
</DictionaryEntry>

```

[9'] Version papier

blonde *n.f.* queb. ragazza, compagna *Ma blonde et moi on est tombé en amour au moment où on s'est vu* *La mia ragazza ed io ci siamo innamorati a prima vista*

Conclusion

Au sujet de l'intégration de la multiplicité linguistique de la francophonie dans le dictionnaire de langue française, deux positions opposées ont été assumées par deux grands noms de la lexicographie française: Alain Rey et Josette Rey-Debove. Alors qu'Alain Rey (1983 : 568) croit que « le dictionnaire de langue, après avoir contribué à l'édification d'une norme en partie fictive [...] peut se permettre aujourd'hui d'accepter la pluralité des usages », Josette Rey-Debove (1998 : 278) qualifiait d'utopie le projet de décrire dans un même dictionnaire de langue toutes les particularités du français, en jugeant que cela ne serait d'aucune utilité pour les usagers francophones. La récente étude de Verreault et Galarneau (2009) a démontré que ces deux positions opposées ne peuvent pas se rencontrer facilement sur le plan de la pratique lexicographique et que cette aspiration reste encore un mirage linguistique.

Cependant, si ces deux positions divergentes sont difficilement conciliables dans la lexicographie monolingue française, elles doivent devenir nécessairement

complémentaires lorsque l'objet en question est la lexicographie bilingue. S'adressant à des usagers étrangers non francophones, potentiellement exposés à toutes les variétés géographiques de la langue française, le dictionnaire bilingue est tenu d'accepter la pluralité des usages.

Dans le nouveau dictionnaire bilingue que nous réalisons, nous tentons de mettre en œuvre une nouvelle vision du rôle et de la fonction du dictionnaire bilingue aujourd'hui. Conçu à l'origine pour le support papier et électronique afin de répondre aux besoins du plus grand nombre d'utilisateurs et basé sur l'exploitation de différentes ressources lexicologiques québécoises consultables en ligne, ce nouveau dictionnaire « bi-langues » vise à restituer une image plus vivante et égalitaire de la langue française, considérée comme un système ouvert et changeant qui, au sein d'une unité, recueille une multiplicité de langues « autres », toutes également actuelles et légitimes.

Bibliographie

- Boulanger (J.-Cl.), 1998 : « La lexicographie québécoise entre Charybde et Scylla : le Dictionnaire CEC jeunesse », dans *International Journal of Lexicography*, vol. 1, p. 127-150.
- Boysen (G.), 1990 : « L'information syntaxique dans les dictionnaires bilingues », dans *Cahiers de lexicologie*, n° 56-57, p. 47-48.
- Dugas (J.-Y.), 1979 : « Les canadianismes au *Petit Robert* 1978 », dans *Meta*, vol. 24, n° 3, p. 395-410.
- Emirikian (L.) et Fouqueré (C.), 2003 : « Présentation: TALN, Web et corpus », dans *Revue québécoise de linguistique*, vol. 32, n° 1, p. 7-9.
- Le Meur (A.) et Derouin (M.-J.), 2006 : « Lemma-oriented dictionaries, concept-oriented terminology and translation memories », dans Calzolari (N.) *et al.*, éd., *LREC 2006 proceedings* (22-28 may 2006), Genoa, Italy, dernière consultation le 5 novembre 2006, www.xmllex.net/lexicography/lexterm/lrec2006.pdf.
- Poirier (Cl.), 2003 : « Variation du français en francophonie et cohérence de la description lexicographique », dans Cormier (M.C.), Francœur (A.) et Boulanger (J.-Cl.), éd., *Les dictionnaires Le Robert. Genèse et évolution*, Montréal, Presses de l'Université de Montréal, p. 189-226.
- Pruvošt (J.), 2002 : « Traduire l'écart culturel dans les dictionnaires bilingues ? », dans Delas (D.), éd., *Traduire 2*, Centre de Recherche Texte Histoire, Paris, Les Belles Lettres, p. 135-158.
- Rey (A.), 1983 : « Norme et dictionnaires (domaine du français) », dans Bédard (E.) et Maurais (J.), éd., *La norme linguistique*, Québec/Paris, Conseil de la langue française-le Robert, p. 541-569.
- Rey-Debove (J.), 1998 : *La linguistique du signe. Une approche sémiotique du langage*, Paris, Armand Colin.
- StatCan, 2009 : *Statistique Canada*, dernière consultation le 14 novembre 2009, www.statcan.gc.ca.
- Verreault (C.) et Galarneau (A.), 2009, « L'inclusion des Particularismes Extra-Hexagonaux dans la Dernière Édition du *Grand Robert*: Réalité ou Mirage de la Francophonie ? », dans *International Journal of Lexicography*, vol. 22, n° 3, p. 259-280.
- Walter (H.), 1988 : *Le Français dans tous les sens*, Paris, Robert Laffont (Le Livre de poche).

Zotti (V.), 2003 : « La lexicographie bilingue contemporaine : bilan et perspectives (domaine italien-français) », dans *Pubblicazioni dell'Università di Bari, Annali della Facoltà di Lingue e Letterature Straniere*, Terza Serie/2002-2003/XVI, Fasano, Schena Editore, p. 379-399.

Zotti (V.), 2007 : *Dictionnaire bilingue et francophonie : le français québécois*, Fasano, Schena Editore, 134 p. (Biblioteca della ricerca).

Dictionnaires

Bergeron (L.), 1981 : *Dictionnaire de la langue québécoise*, Montréal, VLB éditeur.

Boch : *Il Boch. Dizionario francese italiano italiano francese*, 2000, 2007, quatrième et cinquième éditions, Bologne, Zanichelli.

Clapin (S.), 1894 : *Dictionnaire canadien-français ou Lexique-glossaire des mots, expressions et locutions ne se trouvant pas dans les dictionnaires courants et dont l'usage appartient surtout aux Canadiens-français*, Montréal, C. Beauchemin.

Garzanti : *Grande Dizionario Francese Garzanti*, 2000, 2007, Milan, Garzanti.

Paravia : *Il Nuovo Hachette-Paravia dizionario francese italiano italiano francese*, 1999, 2007, première et deuxième éditions, Turin, Paravia.

Le Petit Larousse : *Le Petit Larousse illustré*, 2005, dernière édition consultée, Paris, Larousse.

Poirier (P.), 1993 : *Glossaire acadien*, Moncton, édition d'Acadie, édition critique établie par Pierre M. Gerin.

Le Petit Robert : *Le Petit Robert 1. Dictionnaire alphabétique et analogique de la langue française*, Paris, Le Robert, 1991 ; *Nouveau Petit Robert*, version électronique, Paris, Dictionnaires Le Robert/VUEF, 2001.

DHFQ, 1998 : *Dictionnaire historique du français québécois : monographies lexicographiques de québécoisismes*/Trésor de la langue française au Québec, sous la direction de Claude Poirier, Edith Lessard. — Sainte-Foy, Les Presses de l'Université Laval.

Sources sur Internet

BDLP-Québec, 2005 : *Base de données lexicographiques panfrancophone-Québec*, Trésor de la langue française au Québec, dernière mise à jour le 18 mars 2005, www.tlfq.ulaval.ca/bdlp/quebec.asp.

ILQ, 2009 : *Index lexicologique québécois*, Trésor de la langue française au Québec, dernière mise à jour le 25 septembre 2009, www.tlfq.ulaval.ca/ilq.

Annexes

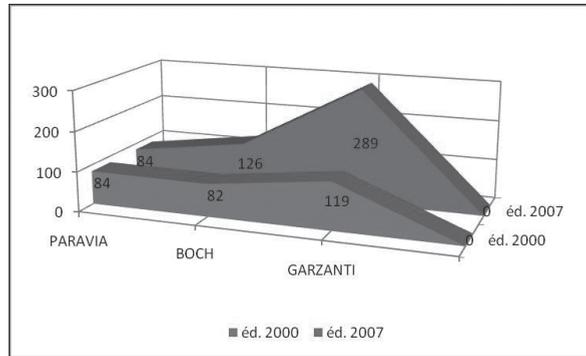


Image 1: Graphique

Formes 1 à 2, sur un total de 2 formes trouvées.

blonde

15 sources, 31 relevés

| Références | [tri chronologique] | Pages |
|--|---------------------|--|
| 1. Barbeau-2: BARBEAU Victor, <i>Le français du Canada</i> , Québec, Garsneau, 1970, 305 p. | | 100' (n.), 201' (n.), 205' (n.), 213' (n.), 227' (n.), 243' (n.), 84' (n.), 87' (n.) |
| 2. Barblan: BARBEAU Victor, <i>Le ramage de mon pays</i> , Montréal, Éditions Bernard Valiquette, 1979, 225 p. | | 175' (n.), 207' (n.), 82' (n.) |
| 3. Bergsuppl 1981: BERGERON Léandre, <i>Outonnaire de la langue québécoise. Supplément 1981, préloché de La charte de la langue québécoise</i> , Montréal, VLB éditeur, 1981, 168 p. | | 113' (n.), 74' (n.), 80' (n.), 95' (n.) |
| 4. Bétière-1: BÉTIÈRE Louis-Alexandre, <i>Dictionnaire général de la langue française au Canada</i> , Québec, Bétière éditeur, 1957, [XIV]-1190 p. | | 183' (n.), 37' (n.) |
| 5. Bétière-3: BÉTIÈRE Louis-Alexandre, <i>Dictionnaire nonl-américain de la langue française</i> , Montréal, Beauchemin, 1975, [XIV]-1194 p. | | 149' (n.), 31' (n.) |
| 6. Clapin: CLAPIN Sylvie, <i>Dictionnaire canadien-français ou lexique-glossaire des mots, expressions et locutions ne se trouvant pas dans les dictionnaires courants et dont l'usage appartient surtout aux Canadiens-français ...</i> , Montréal-Boston, C. O. Beauchemin et fils-Sylvia Clapin, 1994, XLV-389 p. | | 162' (n.) |
| 7. DesChênes: DESCHÊNES Pierre, <i>Le livre des expressions québécoises</i> , Montréal, Hurtubise HMN, 1979, 291 p. | | 64' (n.) |
| 8. Gédéchal: GÉDECHAL James Jr., <i>Study of an Acadian-French Dialect the North Shore of the Bas-de-Chaleurs</i> , Halifax, Max Niemeyer, 1908, XVIII-317 p. | | 170' (n.) |
| 9. Gédéchal-8: GÉDECHAL James Jr., <i>Study of an Acadian Dialect Spoken on the North Shore of the Bay of Chaleur</i> (manuscrit), Harvard University (Cambridge, Mass.), [1914], 204 p. | | 129' (n.) |
| 10. HéonIN: HÉON U. (sère), <i>Quelques mots et expressions en usage sur les îles de la Nouvelle-Bretagne</i> (manuscrit dactylographé), [1941], [12 p.], 9' (n.) | | 1' (n.), 9' (n.) |
| 11. LavChic: LAVOIE Thomas, <i>Le français parlé à Chicoutimi (Canada) (thèse de doctorat)</i> , Faculté des lettres et sciences humaines de Strasbourg, 1970, XXXIV-336 p. | | 193' (n.), 246' (n.) |
| 12. LeClé: LECLERC Paul-André, <i>Les voitures à cheval à la campagne</i> , La Pocatière (Québec), Musée François-Plé, 1970, 129 p. | | 117' (n.) |
| 13. OrkinCan-2: ORKIN Mark H., <i>Speaking Canadian French</i> , 2e édition, Toronto, General Publishing Company Limited, 1971, XII-132 p. | | 67' (n.) |
| 14. OrkFr: ORKIN Mark H., <i>French Canadian, etc?</i> , [Toronto], Lester and Orpen, 1975, 126 p. | | 51' (n.) |
| 15. Pointier: POIRIER Pascal, <i>Glossaire acadien, fasc. 2</i> , Université Saint-Joseph (N.-B.), [1927] 1993, 163 p. | | 84' (n.) |

Image 2: « blonde »

| Fermer | 3 citations | Renvois | Français de référence | Étymologie/Historique |
|--|-------------|--|-----------------------|-----------------------|
| Velette | | Définition | | |
| achalandage 01. [aʃalãdaz]  (n. m.) | | Action de fréquenter en grand nombre un endroit, un lieu, d'y affluer ; résultat de cette action. | | |
| | | Fort, faible achalandage. Hausse,  baisse de l'achalandage. | | |
| | | [État des données: avancé] | | |
| Citations | | Référence | | |
| La circulation [automobile entre les villes de l'Ontario central] y est donc bien moins intense que dans la vallée du Saint-Laurent. [...] ce faible achalandage ne peut s'expliquer par un mauvais état de la route. | | 1965, L.-E. Hamelin, dans CGQ 10/19, p. 44. [études scientifiques] | | |
| Il y en a du va-et-vient près des écuries de la piste de courses de poneys, tant avant qu'après chaque course. L'achalandage va certes diminuer au cours des prochains jours alors que la saison 1980 s'est officiellement complétée dimanche dernier. | | 1980, <i>Peuple-Tribune</i> , Lévis, 10 sept., p. B14. [presse, journaux, périodiques] | | |
| [...] le port de Québec a été témoin de 105 escales de navires de croisière, qui ont amené dans la Vieille Capitale 32,000 visiteurs et 16,000 membres d'équipage. [...] M. Dave Johanson, gérant de la promotion au port de Québec, a attribué cet achalandage à divers facteurs [...]. | | 1987, <i>Le Soleil</i> , 24 oct., p. A9. [presse, journaux, périodiques] | | |

Image 3: « achalandage »

Trésor de la langue française au Québec

[Accueil](#) · [Fonds documentaires](#) · [Thèses et mémoires](#) · [Publications](#) · [Ressources Internet](#) · [Cours](#) · [Plan du site](#)

ILQ - Résultats de recherche

-  Recherche de mots ou d'expressions
-  Recherche de sources
-  Recherche de thèmes
-  Guide de l'utilisateur

Formes 1 à 22, sur un total de 22 formes trouvées.

| | | |
|---|------------------------------------|----------------------------------|
| [branches (apprendre à travers les ...)] | [tort à travers (à ...)] | [traverse à niveau] |
| [branches (à travers les ...)] | [tort et à travers (à ...)] | [traverser les champs à la nage] |
| [branches (entendre à travers les ...)] | [travers (à ...)] | |
| [branches (savoir à travers les ...)] | [travers (de bord à ...)] | |
| [chapeau (à travers son ...)] | [travers (passer à ...)] | |
| [chapeau (parler à travers son ...)] | [travers à l'autre (d'...)] | |
| [as lui passent à travers la peau (les...)] | [travers à l'autre (d'un...)] | |
| [piquer à travers] | [travers de (à ...)] | |
| [piquer à travers champs] | [travers les yeux de (voir à ...)] | |
| [scie à travers] | [traverse (à ...)] | |

travers (passer à ...)

2 sources, 2 relevés

| Références | [tri chronologique] | Pages |
|---|---------------------|-------|
| 1. RobMan-1: ROBINSON Sinclair et SMITH Donald, <i>Practical Handbook of Canadian French. Manuel pratique du français canadien</i> , Toronto, Macmillan of Canada, 1973, 172 p.  | | 171 |
| 2. SimManne: SIMARD Gilbert, <i>La manne bleue</i> , Chicoutimi, Imprimerie Léopold Tremblay Itée, 1978, 129 p.  | | 125 |

Image 4: « passer à travers »

Samia Bouchaddakh,
Centre d'analyse et de traitement informatique du français québécois (Catifq),
Université de Sherbrooke,
Observatoire de linguistique Sens-Texte (OLST), Université de Montréal,
Canada,
samia.bouchaddakh@usherbrooke.ca.

Le Dico-FRAR : base de données lexicographiques bilingue français-arabe

Résumé

Cet article présente le Dico-FRAR, une base de données lexicographiques bilingue français-arabe. Cette base de données a été conçue en vue de proposer à l'utilisateur francophone qui désire s'exprimer en arabe, non seulement la traduction des unités françaises, mais également une description globale des unités lexicales dans les deux langues mises en parallèle. Les équivalents fournis sont décrits de manière systématique sous leurs différentes facettes : sémantique, syntaxique et lexico-combinatoire. L'article comprend trois sections principales : problématique générale et mise en contexte, description de la base Dico-FRAR et exemple de traitement lexicographique bilingue tel qu'il est proposé dans cette base.

Mots-clés : base de données relationnelle ; lexicographie bilingue ; lexicologie explicative et combinatoire ; traduction ; divergences interlinguistiques ; encodage.

1 Introduction : problématique et contexte

Le traitement des vocables proposé dans les dictionnaires bilingues français-arabe n'est pas toujours satisfaisant, notamment pour les usagers francophones qui consultent ces ouvrages en vue de produire un texte dans la langue cible qu'ils ne maîtrisent pas encore parfaitement. En effet, dans une perspective d'encodage, les usagers recourent à ces dictionnaires bilingues dits « de thème » pour connaître certes l'équivalent traductif arabe à une unité lexicale française, mais également toutes les informations sémantiques, syntaxiques ainsi que la combinatoire lexicale qui lui sont rattachées, afin de pouvoir l'employer correctement dans la langue cible. Ces usagers ont aussi besoin d'avoir plus de données sur les divergences lexicales, sémantiques, stylistiques, etc. qui existent entre les deux langues mises en présence, afin de prévenir les éventuelles erreurs et de réussir efficacement leurs tâches d'encodage et de traduction.

Cependant, les nombreuses lacunes et faiblesses dont souffrent les dictionnaires bilingues français-arabe représentent un handicap pour la réussite de telles tâches.

Parmi ces faiblesses, citons la multiplicité des équivalents proposés, le manque de précision dans les données sémantiques et syntaxiques consignées, la non-identification des collocations, l'absence d'informations sur les divergences interlinguistiques qui existent entre les deux langues, le manque de structure dans l'information contenue dans les articles, etc.

L'insatisfaction des usagers de ces dictionnaires est fortement ressentie du fait que la langue arabe, très différente du français, leur est difficilement accessible. En effet, en cas d'absence ou manque d'informations, l'utilisateur ne peut facilement aller chercher un complément d'information dans un dictionnaire unilingue. La consultation d'un dictionnaire arabe n'est pas une tâche facile pour un locuteur non arabophone. Les entrées sont très souvent structurées selon la racine et non d'après la forme du lexème¹. Cette tradition lexicographique arabe, qui pose la dérivation comme principe d'organisation des entrées dans le dictionnaire, suppose un savoir préalable et pose problème pour les locuteurs non arabophones.

Par ailleurs, l'on assiste depuis plusieurs années au développement d'une variété de bases de données lexicales dans différentes langues, ce qui a permis de réaliser des avancées notables dans le domaine lexicographique unilingue – par exemple, WordNet (Fellbaum 1998), FrameNet (Baker *et al.* 1998, Fillmore *et al.* 2001), DiCo (Polguère 2000, 2003), Dafles (Selva *et al.* 2002, Verlinde *et al.* 2005). En revanche, en lexicographie bilingue, c'est au cours de la dernière décennie que des travaux intéressants ont commencé à se multiplier – par exemple, Fontenelle (1997), Alonso Ramos (2001). Toutefois, la majorité de ces travaux sont concentrés, notamment, sur le traitement des langues indo-européennes, comme l'anglais et le français. Dans la paire de langues français-arabe, en particulier, il n'existe pas encore, à notre connaissance, de ressource lexicographique satisfaisante susceptible d'être mise à la disposition des traducteurs et des apprenants de l'arabe langue étrangère.

Dans ce contexte, nous avons examiné les problèmes qui se posent dans les dictionnaires bilingues français-arabe et développé une base de données lexicales visant à améliorer les descriptions qu'on y trouve, particulièrement dans ceux dits « de thème » ou « d'encodage » (Bouchaddakh 2010, en préparation). Cette base de données (BD), que nous proposons de présenter dans les pages qui suivent, est baptisée « Dico-FRAR ».

Notre exposé est organisé de la façon suivante. Nous présentons d'abord les propriétés générales de notre BD (section 2.1). Ensuite, nous définissons le contenu informationnel que renferment nos articles de dictionnaire bilingues (section 2.2), avant d'introduire, dans ses grandes lignes, la structure générale de la BD réalisée

¹ Tous les lexèmes, excepté quelques unités grammaticales, sont ramenés à une racine généralement trilitère représentant un contenu sémantique commun. Les lexèmes dérivés à partir de cette racine, dont la forme est purement consonantique, sont inclus au sein de cet article de dictionnaire. Nous pouvons citer, pour illustration, l'exemple canonique de la racine KTB qui donne par dérivation le verbe [kataba] « écrire », [kitâb] « livre », [kuttâb] « école islamique », [kâtib] « écrivain », [maktûb] « écrit ou lettre », etc. Toutes ces unités sont décrites sous l'entrée KTB.

et de montrer son intérêt, aussi bien pour l'utilisateur que pour le lexicographe (section 2.3). Finalement, dans la troisième section, nous donnons un exemple d'article bilingue français-arabe, ce qui permet d'illustrer, par le fait même, le traitement que nous préconisons dans une base de données lexicographiques bilingue.

2 Dico-FRAR

2.1 Caractéristiques générales

La base de données lexicographiques Dico-FRAR repose sur les principes de la lexicologie explicative et combinatoire (LEC) (Mel'čuk *et al.* 1995) dont le modèle lexicographique permet de combler, à plus d'un égard, les lacunes et les faiblesses des dictionnaires bilingues traditionnels.

Conçue principalement pour l'apprenant francophone de la langue arabe qui est de niveau intermédiaire ou avancé, ainsi que pour le traducteur et pour tous ceux qui s'intéressent à la production d'énoncés en arabe, la base Dico-FRAR est composée des trois parties suivantes. La première est réservée à la description unilingue des unités françaises, la deuxième prend en charge la description des unités lexicales arabes et la troisième a pour objet la mise en relation de ces unités françaises et arabes dans une optique d'encodage.

L'unité de description dans cette base est la lexie². Une fois que les différentes acceptions du vocable ont été distinguées et ordonnées, chaque lexie est décrite sur les plans sémantique, syntaxique et lexico-combinatoire avec une illustration de son emploi dans différents exemples.

Pour les unités arabes, les articles de dictionnaire sont élaborés dans un métalangage français, ce qui rend l'accès à l'information plus aisé pour un usager francophone et facilite aussi la mise en correspondance des deux langues (le français et l'arabe), ainsi que la sélection des meilleurs équivalents. Dans Bouchaddakh (2008), nous avons montré que la structuration explicite des définitions rédigées dans un même métalangage, selon les principes de la LEC, permettait de sélectionner adéquatement les équivalents à proposer dans un dictionnaire bilingue. L'article bilingue, qui résulte de cette mise en correspondance, fournit à l'utilisateur toutes les informations qui lui sont nécessaires dans sa tâche d'encodage ou de traduction. Il propose ainsi à ce dernier des équivalents plus appropriés, souligne les divergences interlinguistiques (sémantiques, stylistiques, syntaxiques, etc.) entre deux lexies dites « équivalentes » et, enfin, présente des solutions en cas d'absence d'équivalent et des mises en garde contre les éventuelles erreurs d'encodage. Cela étant, passons maintenant à la présentation du contenu informationnel d'un tel article bilingue.

² Dans la terminologie de la LEC, une lexie est soit un lexème soit une locution associé(e) à un sens donné. Un regroupement de lexies qui ont le même signifiant et un lien sémantique évident est appelé un « vocable ».

2.2 Contenu informationnel de l'article bilingue

Afin d'offrir un traitement lexicographique rigoureux, l'article bilingue, qui est ainsi la synthèse de la mise en parallèle des descriptions des unités lexicales françaises et arabes, doit renfermer un certain nombre d'informations que nous définissons comme suit.

2.2.1 Informations liées à l'identification de la lexie-vedette

- L'entrée lexicale qui est le nom du vocable français.
- Le numéro du vocable français pour distinguer les homonymes s'il y a lieu.
- La partie du discours (désormais PDD) du vocable français.
- Le numéro d'acceptation de la lexie à traduire. Ce numéro est identique à celui qui a été proposé dans l'article unilingue français de la même lexie. Par souci d'uniformité, nous gardons dans l'article bilingue le même dégroupement des sens des vocables que celui proposé dans nos articles unilingues. La consultation des différents articles n'en sera que facilitée pour l'utilisateur.
- La forme propositionnelle (ou structure actancielle) de la lexie française qui permet d'explicitier le nombre de ses actants. Par exemple, pour la lexie *châtier* (cf. section 3), la forme propositionnelle est *X châtie Y pour Z par W*. On pourrait également caractériser (ou « typer » dans la terminologie de la LEC) les différentes variables X, Y, etc. dans la forme propositionnelle, ce qui nous donnerait quelque chose comme : *l'individu X châtie l'individu Y pour l'acte Z par le mal W*.

2.2.2 Informations liées à l'identification de l'équivalent principal proposé

- L'équivalent traductif arabe qui est, bien entendu, une lexie et non un vocable.
- La racine de l'unité arabe, s'il y a lieu. La notion de « racine » revêt, en effet, une importance particulière dans l'apprentissage de la langue arabe.
- Un *nota bene* pour préciser la nature de l'équivalent fourni (comme le montre l'article proposé dans la section 3) ou pour renvoyer l'utilisateur à une variante ou à un autre équivalent possible, selon le contexte. C'est également dans cette section qu'on attire l'attention de l'utilisateur sur la forme du féminin ou du pluriel non régulière pour les noms et les adjectifs.
- La PDD de l'équivalent. Généralement cette PDD correspond à celle de la lexie française, mais l'équivalent peut parfois appartenir à une PDD différente ou particulière à la langue arabe.
- La forme propositionnelle de la lexie arabe.

2.2.3 Informations sémantiques

Les informations d'ordre sémantique sont d'une grande utilité pour l'utilisateur, d'autant plus que la majorité des équivalents traductifs proposés sont des quasi-équivalents ou des équivalents partiels. Ces informations seront classées dans deux sections

séparées. La première reprend la définition de la lexie telle qu'elle apparaît dans l'article unilingue arabe. La deuxième porte sur la ou les divergences sémantiques entre les deux lexies, française et arabe, utiles pour que l'utilisateur prenne réellement conscience du sens véhiculé par tel ou tel équivalent et puisse décider, avec discernement, de sa pertinence ou non dans un contexte donné.

2.2.4 Informations liées à la combinatoire syntaxique

- Une description du régime syntaxique de l'équivalent.
- Une mise en garde contre les éventuelles erreurs dues aux divergences, dans le comportement syntaxique, entre la lexie française et son équivalent arabe. Par exemple, les alternances syntaxiques peuvent être différentes d'une langue à une autre. C'est le cas de *châtierI* et de son équivalent arabe (cf. section 3).

2.2.5 Informations liées à la combinatoire lexicale

Ces informations sont de deux types :

- Le premier type concerne les lexies reliées à l'équivalent par des liens paradigmatiques. Ces liens relèvent de la dérivation sémantique comme les synonymes, les converses, les antonymes, ainsi que les différents noms actanciels typiques de la lexie, la contrepartie verbale d'une lexie nominale ou, au contraire, la contrepartie nominale d'un verbe ou d'un adjectif, etc.
- Le deuxième type porte sur les lexies reliées à l'équivalent par des liens syntagmatiques. Il est ici question, bien entendu, des collocations – expressions semi-figées bipartites composées de deux éléments : la base (la lexie-vedette) et le collocatif (Hausmann 1979, Mel'čuk 1996). Contrairement au premier type d'informations lexicales, il ne s'agit pas, dans ce cas-ci, de répertorier simplement les lexies qui sont en lien syntagmatique privilégié avec la lexie arabe (les collocatifs), mais également de préciser leur sens et leur emploi, et de proposer des équivalents traductifs aux collocations françaises, le tout illustré d'exemples traduits en français.

Dans l'approche de la LEC, la modélisation de ces deux types de relations est réalisée au moyen des fonctions lexicales (FL). Faute d'espace, nous ne nous attardons pas, ici, à la description du mécanisme qui les sous-tend ; celui-ci est, à notre avis, suffisamment connu (Mel'čuk 1996, Mel'čuk *et al.* 1995 et Mel'čuk et Polguère 2007). Disons simplement que les FL spécifient l'ensemble des possibilités de combinatoire paradigmatiques (les dérivations sémantiques) et syntagmatiques (les collocations ou les co-occurrences lexicales restreintes) d'une unité lexicale donnée. Par exemple, pour *remercier*, la FL So, qui renvoie à la nominalisation du verbe, a comme valeur *remerciement* et la FL Magn, qui dénote l'intensification, a comme valeur *énormément* et *infiniment*. Notons que, dans nos articles de dictionnaire, les FL qui sont jugées difficiles à saisir pour un usager non familiarisé avec ce type d'encodage, sont vulgarisées par des formules qui en explicitent le sens. Par exemple, la FL Magn est explicitée par *De façon sévère* dans notre description de la lexie *châtier* (section 3).

2.2.6 Illustration par des exemples

Les exemples, dans l'article bilingue, sont deux sortes :

- Des exemples illustrant l'emploi des équivalents proposés dans la langue cible. Ceux-ci sont traduits en français, souvent littéralement, pour mieux aider l'utilisateur à saisir l'emploi des unités arabes.
- Un ou deux exemples de traduction avec l'équivalent proposé. Contrairement aux premiers, il s'agit ici d'exemples français traduits en arabe avec l'emploi de l'équivalent fourni.

Les figures 2 et 4 présentées dans la section 2.3.2 montrent de façon concrète comment est organisé le contenu informationnel de notre article bilingue. Il est à noter que celui-ci concerne surtout l'article bilingue avec équivalent direct en arabe, c'est-à-dire avec un équivalent qui est une lexie arabe décrite sur les plans sémantique, syntaxique et lexical. Les lexies françaises qui n'ont pas d'équivalent direct en arabe, mais correspondent plutôt à des collocations ou à des expressions libres dans la langue cible, sont décrites dans un deuxième type d'article bilingue, simplifié, prévu à cet effet.

En résumé, notre BD comporte quatre types d'articles : des articles unilingues français, des articles unilingues arabes, des articles bilingues avec équivalent direct et des articles bilingues sans équivalent direct. Dans la section suivante, nous verrons comment cette BD est structurée.

2.3 Structure du Dico-FRAR

Le Dico-FRAR est une base de données dictionnaire. Chacun des articles qu'elle contient présente les données sous forme de texte. Cela dit, le texte lexicographique n'a pas besoin d'être compilé en un seul bloc. Avec l'outil informatique, il est possible de structurer l'article de dictionnaire sous différents blocs plus ou moins indépendants, ce qui facilite la consultation pour l'utilisateur et la gestion des données pour le lexicographe. L'outil informatique permet, par exemple, d'afficher plus aisément les relations sémantiques et lexicales entre les diverses unités décrites et de réviser rigoureusement le contenu informationnel fourni.

Ainsi, pour concevoir notre base de données français-arabe, notre choix s'est arrêté sur une base relationnelle³ gérée par le logiciel *Access* de Microsoft Office. Pour chacun des quatre types d'articles que renferme notre BD, nous avons conçu des tables pour contenir les diverses données. Cette façon de faire, c'est-à-dire de créer des tables séparées, permet de gérer facilement la base et de l'enrichir, au besoin, en ajoutant des données ou des champs, sans avoir à la réaménager à chaque fois. Au total, notre BD est structurée en seize tables distinctes dont certaines sont reliées

³ Une base de données relationnelle est une base dans laquelle les données sont stockées dans des tables distinctes, selon le type d'information à enregistrer, mais qui sont reliées entre elles et peuvent être regroupées selon les besoins.

entre elles pour former nos quatre types d'articles. Nous nous limitons, dans ce texte, à la description de la structure de l'article bilingue avec équivalent direct.

2.3.1 Organisation des tables de l'article bilingue avec équivalent direct

Pour cet article dictionnaire, nous avons créé quatre tables réservées chacune à un type d'information particulier, à savoir: *Entree_bilingue_ED*⁴, *Relations_paradigmatiques_ar*, *Collocations_fr_ar* et *Vulgarisations_fl*.

La première, *Entree_bilingue_ED*, est la table principale de l'article bilingue avec équivalent direct. Les différents champs qu'elle renferme présentent les données caractéristiques essentielles de la lexie-vedette française à traduire et de son équivalent arabe comme le nom du vocable, le numéro de la lexie, la PDD, etc. Ces données sont directement importées des tables liées aux articles unilingues français et arabes. La table contient également les champs relatifs aux données sémantiques et syntaxiques associés à la lexie arabe, ainsi qu'aux exemples qui illustrent l'emploi de cette lexie dans un contexte de la langue cible.

La table *Relations_paradigmatiques_ar* renferme, comme son nom l'indique, les informations liées aux relations lexicales paradigmatiques de l'équivalent arabe fourni. Ces informations étant également déjà stockées dans une des tables de l'article unilingue arabe, nous les avons simplement sélectionnées et importées directement depuis la table en question.

La table *Collocations_fr_ar* est destinée à contenir les données relatives aux relations lexicales syntagmatiques de la lexie française et de son équivalent arabe.

Enfin, la table *Vulgarisations_fl* a été créée pour nous permettre de regrouper différentes valeurs d'une même fonction lexicale donnée.

Pour mieux expliquer l'utilité de cette dernière table, rappelons que l'application d'une seule fonction lexicale ou d'une seule vulgarisation de fonction lexicale à une lexie peut donner plusieurs valeurs qui sont plus ou moins différentes sur le plan sémantique. Prenons, à titre d'exemple, la lexie *punition* dont la forme propositionnelle est *punition imposée par X à Y pour l'acte de Z et consistant en W*. Les collocations suivantes *punition méritée/juste/équitable* sont toutes encodées, dans le formalisme de la LEC, avec la même fonction lexicale *Ver_(punition)* ou sa vulgarisation *Que l'acte Z justifie*, mais séparées avec un point-virgule. Toutefois, dans un article bilingue, il importe, selon nous, de distinguer plus clairement ces collocations pour proposer à l'utilisateur les bons équivalents qu'il pourra employer sans équivoque. C'est pourquoi, bien qu'elles soient décrites avec la même fonction lexicale ou la même vulgarisation de fonction lexicale, nous avons départagé ces collocatifs en les citant comme des enregistrements distincts dans la table *Collocations_fr_ar*

4 Pour éviter tout problème avec *Access*, les accents n'ont pas été utilisés dans les noms des tables et des champs créés.

(voir figure 1). Avec la table *Vulgarisations_fl*, nous pouvons faire cette distinction sans perdre de vue le lien entre les différentes données.

| Nom_vocable_fr | Num_lexie_fr | Nom_vocable_ar | Num_lexie_ar | Vulgarisation_fl | Collocatifs_arabes | Collocatifs_français |
|----------------|--------------|-------------------|--------------|-----------------------|----------------------|----------------------|
| PUNITION | 1 | عَذَاب [ʕiqâb] | 0 | Que l'acte Z justifie | عادل [ʕādil] | équitable, juste |
| PUNITION | 1 | عَذَاب [ʕiqâb] | 0 | Que l'acte Z justifie | مستحق [mustahhaq] | méritée |

Figure 1 : Extrait de la table *Collocations_fr_ar*⁵

Les diverses tables présentées ci-dessus sont reliées entre elles via les champs communs qu'elles contiennent, ce qui permet de garder l'unité de l'article dictionnaire et la cohérence des données qu'il renferme.

2.3.2 La mise en page

Après avoir identifié le contenu informationnel de nos différents articles et structuré à partir de là notre BD en diverses tables reliées entre elles, nous avons conçu un formulaire de saisie pour chacun de nos quatre types d'article. Ce formulaire permet d'enregistrer nos données de façon conviviale, en ayant toujours sous les yeux les informations qui sont réparties de manière automatique dans les différentes tables sous-jacentes. Le formulaire permet aussi de corriger ou modifier efficacement, s'il y a lieu, les données consignées sans avoir à ouvrir nécessairement toutes les tables où des modifications doivent être apportées, au risque d'en oublier quelques-unes. La figure 2 donne un aperçu de ce formulaire pour l'article bilingue avec équivalent direct.

5 Pour faciliter la lecture des unités lexicales arabes au lecteur de cet article, nous utilisons le système de translittération de l'arabe DIN-31635. Les lexèmes translittérés sont mis en italique entre crochets.

| | | | |
|-----------------|-----------|-------|-------------------------|
| Nom_Vocabulaire | Num_Lexie | Pdd | Fp |
| CHÂTIER | 1 | Verbe | X châtie Y pour Z par W |

| | | | |
|-----------------|-----------------|--------|---|
| Nom_Vocabulaire | كَب | Racine | كَب |
| Num_Lexie | 2 | NB | L'équivalent proposé n'est pas un équivalent parfait. Voir la définition et la mise en garde. |
| Pdd | Verbe فعل | | |
| Fp | W ب Z ل Y X كَب | | |

Informations sémantiques

Def
 L'individu X tenant l'individu Y responsable de l'acte Z qui va à l'encontre de certains principes auxquels X se réfère // X, qui a une autorité sur Y, fait subir un mal W à Y afin que Y ne refasse pas Z et améliore sa conduite en respectant les principes de X.

Mise en garde
 Prendre note que la composante sémantique liée à l'amélioration de la conduite de Y est plus forte avec la lexie arabe qu'avec la lexie française. De plus, le mal W ou la punition infligée avec "châtier" est sévère, ce qui n'est pas forcément le cas avec كَب. C'est pourquoi,

Informations syntaxiques

Tr
 X = I = N nominatif
 Y = II = N accusatif oblig.
 Z = III = N ط N ل
 W = IV = N ب

Mise en garde
 Contrairement à "châtier", كَب a un seul régime syntaxique. La présence de l'actant Y (1er complément) est obligatoire dans un énoncé.

Exp.

كَبه المعلم على تصرفه القوي تلميذا عارفا
 كَبها الخروجها المتكرر دون إذن
 كَب الأب أولاده بالمشرب

L'instituteur l'a sévèrement châtié pour son mauvais comportement.
 Il l'a châtié pour ses fréquentes sorties sans permission.
 Le père a châtié ses enfants avec une punition corporelle.

Informations lexicales

Dérivés sémantiques

| Nom_Fl | Vulgarisation | Valeur |
|--------|---------------|-------------|
| A0 | | تأديب |
| A2 | [Y] que X ~ | تمت التأديب |

Enr : 1 sur 8

Collocations

| Vulgarisation de FL | De façon justifiée |
|----------------------|--------------------|
| Collocatifs français | Collocatifs arabes |
| bien, justement | 0 |

Enr : 1 sur 1

Remarques et commentaires
 Ces collocatifs adverbiaux n'ont pas d'équivalents directs en arabe. On peut employer le verbe كَب avec une collocation composée de l'adjectif

Exemples
 كَبه المعلم احسن تلميذا
 Lit. L'instituteur l'a châtié d'un meilleur châtiement

Enr : 2 sur 2

Exemple de traduction

1- On ne doit pas châtier les enfants avant qu'ils puissent connaître leurs fautes.
 - من المفروض أن لا تُلَبِّب الأطفال قبل أن يتعلموا أخطائهم.

2- L'inceste est universellement prohibé et même sévèrement châtié dans toutes les sociétés.
 - يشهد العالم كله وحظور و جرمية تأديب عالياً بتلك في جميع المجتمعات.

Figure 2 : Formulaire de saisie de l'article bilingue avec équivalent direct

Comme nous l'avons signalé ci-dessus, certaines données de l'article bilingue sont directement importées des tables relatives aux articles unilingues et ce, après la mise en équivalence des lexies françaises et arabes.

Pour procéder à cette mise en équivalence des unités lexicales des deux langues, nous avons créé un deuxième type de formulaire que nous avons appelé *formulaire de mise en correspondance des lexies fr-ar*. Celui-ci se présente comme suit :

Figure 3 : Formulaire de mise en correspondance des lexies fr-ar

Avec ce formulaire, qui comprend une partie réservée à l'entrée française et une autre à l'entrée arabe, nous avons accès de manière parallèle aux diverses lexies décrites dans les deux langues. Pour bien identifier les lexies en cause, le formulaire indique le nom du vocable, son numéro, sa PDD, le numéro de la lexie et sa forme propositionnelle. Pour obtenir de plus amples détails sur la lexie française ou arabe choisie, le formulaire offre en outre un accès direct aux articles unilingues (par le biais des onglets *Consultation dictionnaire unilingue français* et *Consultation dictionnaire unilingue arabe* en haut du formulaire dans la figure 3).

Le formulaire contient aussi d'autres types d'informations dont nous avons besoin pour établir la mise en correspondance entre les lexies françaises et arabes. Nous avons ainsi, à côté du numéro de la lexie, le champ *Sélection* présenté ici sous forme de case dans laquelle on coche pour sélectionner les lexies françaises et arabes équivalentes. Il y a aussi le champ *Ajout* (noté *Entrée inscrite dans le dictionnaire bilingue* dans la fig. 3), présenté également sous forme de case, qui sera cochée automatiquement lorsque la lexie française sélectionnée sera ajoutée dans le dictionnaire bilingue, ce qui permet de distinguer les lexies traitées de celles qui ne le sont pas encore⁶. Cette modification automatique dans le champ *Ajout* est effectuée via une requête de *Mise à jour* que l'on exécute avec un des deux boutons de commande nommés *Mettre à jour le dictionnaire bilingue* (selon que le type d'article bilingue choisi est avec ou sans équivalent direct). Une fois que nos

6 Rappelons que notre traitement lexicographique bilingue est unidirectionnel du français vers l'arabe. Donc, une lexie française doit être traitée une seule fois. Par contre, une lexie arabe peut être proposée comme équivalent possible à plusieurs lexies françaises.

deux lexies équivalentes ont été sélectionnées et ajoutées à la table *Entree_bilingue* correspondante, deux autres boutons de commande permettent d'accéder au formulaire de saisie de l'article bilingue lié à cette nouvelle entrée pour compléter la rédaction de l'article (voir en bas du formulaire dans la figure 3).

Les formulaires présentés plus haut sont, comme leur nom l'indique, des formulaires de saisie et de mise en correspondance. Ils sont donc destinés au lexicographe/rédacteur et non à l'utilisateur (l'apprenant ou le traducteur) de la BD. Par conséquent, pour une consultation des articles sans risque de modification des données enregistrées, des pages web ont été créées. Ces pages permettent d'avoir accès aux fiches lexicographiques contenues dans la base, via le web ou l'intranet.

Pour chaque type d'article de dictionnaire (unilingue arabe, unilingue français, bilingue français-arabe), nous avons conçu diverses pages d'accès aux données, à savoir une page d'accès pour l'article unilingue arabe, une autre pour l'article unilingue français, deux pages pour l'article bilingue avec équivalent direct (version complète et version abrégée) et, finalement, une page pour l'article bilingue sans équivalent direct. Toutes ces pages peuvent être consultées à partir d'une page d'accueil qui sert de point d'entrée à Dico-FRAR. L'utilisateur peut ainsi choisir, selon ses besoins, l'un ou l'autre des types d'article de dictionnaire auquel il veut accéder, ainsi que le niveau de détail souhaité pour l'affichage des données. Celles-ci, classées en diverses rubriques, ne s'affichent que si l'utilisateur le désire.

Les figures suivantes présentent un exemple de la page d'accès de l'article bilingue avec équivalent direct dans sa version abrégée (fig. 4) et sa version complète (fig. 5). Le contenu de cet article est présenté en détail à la section 3.

The screenshot displays the 'CHÂTIER' entry in the Dico-FRAR system. It shows the French form 'CHÂTIER' (Verb) with a frequency of 0. Below this, it provides the French form 'X châtie Y pour Z par W' and the Arabic equivalent 'On craint d'être châtiés pour avoir commis ce péché.' The interface includes a search bar for Arabic terms, showing 'أُتِب' with a frequency of 0 and the root 'أُتِب' with a frequency of 2. A note indicates that the proposed equivalent is not perfect. The 'Information lexicale : collocations' section lists 'Vulgarisation de fl.' as 'De façon justifiée'. The 'Collocatifs français' are 'bien, justement', and the 'Collocatifs arabes' are listed as 0. The 'Remarques et commentaires' section explains that these adverbial collocations do not have direct equivalents in Arabic and provides examples of their use. The 'Exemples' section shows the Arabic phrase 'أذبه المعلم أحسن تأديب' (The teacher scolded him better than a scolding) and its French translation 'L'instigateur fa châtié d'un meilleur châtiement'. The 'Exemple de traduction' section provides two examples: 'On ne doit pas châtier les enfants avant qu'ils puissent connaître leurs fautes.' and 'L'inceste est universellement prohibé et même sévèrement châtié dans toutes les sociétés.'

Figure 4 : Page d'accès à l'article bilingue avec équivalent direct. Version abrégée

Nom_vocable_fr: Num_vocable_fr: Pdd_fr

CHÂTIER 0 Verbe

Num_lexie_fr: Fp_fr: Breve_def_fr: Ex_fr

1 X châtie Y pour Z par W X... fait subir un mal sévère à Y afin que Y ne refasse pas Z et améliore sa conduite en respectant les principes de X. On craint d'être châtiés pour avoir commis ce péché.

Équivalent arabe

Nom_vocable_ar: آذَب Racine: آذَب

Num_vocable_ar: 0 Nb: L'équivalent proposé n'est pas un équivalent parfait. Voir la définition et la mise en garde.

Num_lexie_ar: 2 Pdd_ar: 'Verbe' فعل Fp_ar: آذَب ي آذِب آذَب Z ي W

Informations sémantiques

Def_ar: L'individu X tenant l'individu Y responsable de l'acte Z qui va à l'encontre de certains principes auxquels X se réfère | X, qui a une autorité sur Y, fait subir un mal W à Y afin que Y ne refasse pas Z et améliore sa conduite en respectant les principes de X.

Mise_en_garde_1: Prendre note que la composante sémantique liée à l'amélioration de la conduite de Y est plus forte avec la lexie arabe qu'avec la lexie française. De plus, le mal W ou la punition infligée avec "châtier," est sévère, ce qui n'est pas forcément le cas avec آذَب. C'est pourquoi, dans les contextes où châtier est employé pour punir un coupable, un criminel pour un méfait grave, il est préférable d'employer le verbe عاقب qui a un sens plus général et qui ne vise pas l'amélioration de la conduite de

Informations syntaxiques

Tr_ar: X = I = N nominatif
Y = II = N accusatif oblig.
Z = III = N على, N إن
W = IV = N ب

Mise_en_garde_2: Contrairement à "châtier", آذَب a un seul régime syntaxique. La présence de l'actant Y (1er complément) est obligatoire dans un énoncé.

Ex: آذبه المعلم على تصرفه السيئ تأديبا صارما. L'instituteur l'a sévèrement châtié pour son mauvais comportement.
آذبها لتفويتها المتكرر دون إذن. Il l'a châtié pour ses fréquentes sorties sans permission.
آذب الأب أولاده بالضرب. Le père a châtié ses enfants avec une punition corporelle.

Informations lexicales : dérivés sémantiques

Nom_fl_ar: QSym Vulgarisation_fl_ar: Valeur_fl_ar: عاقب punir

Informations lexicales : collocations

Vulgarisation_de_FL: De façon sévère

| Collocatifs_français | Collocatifs_arabes | Remarques_et_commentaires | Exemples |
|--|--------------------|---|---|
| durement, lourdement, rigoureusement, sévèrement | 0 | Ces collocatifs adverbiaux n'ont pas d'équivalents directs en arabe. On peut employer le verbe آذَب soit avec une expression libre composée de la préposition ب + le masdar عاقب. | آذبه المعلم على تصرفه السيئ تأديبا صارما Lit. L'instituteur l'a châtié d'un châtiement sévère pour son mauvais |

Exemple de traduction

1- On ne doit pas châtier les enfants avant qu'ils puissent connaître leurs fautes. - من المفروض أن لا يُؤذَب الأطفال قبل أن يتبين لهم أخطائهم .
2- L'inceste est universellement prohibé et même sévèrement châtié dans toutes les sociétés. - عُشْيَانُ الفِتْنَانِ مُحْتَظَرٌ وَجَزَاءُهُ يُعَاقَبُ عَلَيْهَا بِشَدَّةٍ فِي جَمِيعِ المَجْتَمَعَاتِ.

Figure 5 : Page d'accès à l'article bilingue avec équivalent direct. Version complète

La version détaillée présente l'article bilingue complet avec toutes les informations sémantiques, syntaxiques et lexicales. La version abrégée présente en gros

l'équivalent arabe, sa PDD, sa forme propositionnelle, les collocations arabes de l'équivalent, ainsi que les équivalents des collocations françaises et l'exemple de traduction.

3 Exemple d'article de dictionnaire bilingue dans le Dico-FRAR

L'exemple d'article de dictionnaire bilingue avec équivalent direct qui figure en annexe permettra au lecteur de se faire une idée concrète du traitement lexicographique que nous proposons dans le Dico-FRAR et de la pertinence des informations qui y sont données⁷. La lexie-vedette sélectionnée pour cet exemple est *châtieri*.

Conclusion

Dans cet article, nous avons présenté la base de données lexicographiques bilingue Dico-FRAR qui peut être proposée comme modèle pour le développement d'un dictionnaire informatisé français-arabe d'encodage. La finalité de cette base de données est multiple. Elle offre à l'utilisateur un traitement lexicographique qui répond, autant que faire se peut, à ses attentes en matière d'encodage ou de traduction. Elle propose également à celui-ci la possibilité de consulter non seulement les articles bilingues français-arabe, mais aussi les articles unilingues français et les articles unilingues arabes rédigés dans un métalangage français. De plus, le Dico-FRAR offre au lexicographe un environnement de travail agréable et convivial pour qu'il arrive à consulter ces articles unilingues, à comparer et à trier les données afin de rédiger des articles bilingues satisfaisants. Enfin, avec la variété et la richesse des données sémantiques, syntaxiques et lexicales des vocables qui seront décrits de façon complète dans les deux langues mises en parallèle, cette base de données, une fois enrichie avec un nombre élevé de vocables, pourra, en outre, être exploitée à des fins de recherche dans divers domaines. Nous pensons particulièrement aux domaines de la lexicographie, de la linguistique comparée, voire de l'enseignement et l'apprentissage d'une langue étrangère.

⁷ Comme il a été mentionné à la note 5, le recours à la translittération de l'arabe dans notre exemple d'article de dictionnaire n'a pour objectif que d'aider le lecteur à lire le présent travail.

Châtier
Verbe

Forme propositionnelle

X châtie Y pour Z par W

2أب [addaba]

'Verbe' فعل

Racine

أب [أب]

NB

L'équivalent proposé n'est pas un équivalent parfait. Voir la définition et la mise en garde ci-dessous.

Forme propositionnelle

W أب Y X ل

Informations sémantiques

Définition

L'individu X tenant l'individu Y responsable de l'acte Z qui va à l'encontre de certains principes auxquels X se réfère || X, qui a une autorité sur Y, fait subir un mal W à Y afin que Y ne refasse pas Z et améliore sa conduite en respectant les principes de X.

Mise en garde

Prendre note que la composante sémantique liée à 'l'amélioration de la conduite de Y' est plus forte avec la lexie arabe qu'avec la lexie française. De plus, le mal W ou la punition infligée avec « châtier » est sévère, ce qui n'est pas forcément le cas avec 2أب. C'est pourquoi, dans les contextes où « châtier » est employé pour punir un coupable, un criminel pour un méfait grave, il est préférable d'employer le verbe عاقب, qui a un sens plus général et qui ne vise pas l'amélioration de la conduite de Y. Voir l'exemple de traduction 2.

Informations syntaxiques

Tableau de régime

X = I = N nominatif

Mise en garde

Contrairement à « châtier », 2أب a un seul régime syntaxique. La présence de

l'actant Y (1er complément) est obligatoire dans un énoncé.

Y = II = N accusatif oblig.

Z = III = N على [ʿalā] N ل [l]

W = IV = N ب [bi]

Exemples

'L'instituteur l'a sévèrement châtié pour son mauvais comportement.'

أُتِبِه المَعْلَمُ عَلَى تَصَرُّفِهِ السَّيِّئِ تَأْدِيبًا صَارِمًا.

[ʿaddaba-hu al-muʿallimu ʿalâ taṣarruffi-hi as-sayyiʿ]

ta tîban ṣāʾiriman.]

'Il l'a châtié pour ses fréquentes sorties sans permission.'

أَتْبَاهَا لِحُرُوجِهَا الْمَتَكَرِّرِ دُونَ إِذْنِ.

[ʿaddaba-hâ li-hurûji-hâ al-mutakarriʾi dûna ʿidn.]

'Le père a châtié ses enfants avec une punition corporelle.'

أَتَّبِ الْآبُ أَوْ لَأَدَهُ بِالضَّرْبِ.

[ʿaddaba al-ʾabu ʾawlâda-hu bi-aq-ḍarbi.]

Informations lexicales

Dérivés sémantiques

| Nom_FI | Vulgarisation_FI | Valeur_FI |
|----------------------|------------------|--|
| عقِبَ a | | 'punir' |
| [ʿāqaba] | | |
| أَقْبَصَ [من Y=N] | | 'infliger à qqn un mal identique ou égal à celui causé' [Y=N] |
| [ʾiqtāṣa min N=Y] | | |
| جَزَى [Y=N] | | Lit. 'donner ce qu'une personne mérite, donner son dû à qqn' [Y=N] |
| [jazā] | | |
| سَامَحَ [عن Y=N] | | 'pardonner' [Y=N] |
| [sāmaḥa, ṣafaha ʿan] | | |
| عَا [عن Y=N] | | 'récompenser' [Y=N] |
| [ʿafā ʿan, ḡafara l] | | |

QAnti

| | |
|--|---|
| | [<i>'alāba, kāfa 'ā</i>] |
| | تأديب 2 [<i>ta'dīb</i>] |
| | خطا [<i>ḥaṭā</i>] 'faute' ذنب [<i>ḍanb</i>] 'mal, méfait' |
| | مخالفة [<i>muhālafā</i>] 'infractio(n)' |
| | عقاب, عقوبة [<i>iqāb, 'uqūba</i>] 'peine, punition' |
| | تأديبي [<i>ta'dībī</i>] |
| | تحت التأديب [<i>tahta at-ta'dīb</i>] |
| | يؤدب [<i>yu'addab</i>] |
| | مخطئ [<i>muhīṭ</i>] 'fauteur' |
| | مخالف : [<i>muhāliṭ</i>] 'contrevenant' |
| | جاني, مذنب [<i>jānī, muḍnīb</i>] 'coupable' |

Collocations

| Vulgarisation_FI | Collocatifs français | Collocatifs arabes | Remarques et commentaires | Exemples |
|------------------|--|--------------------|--|---|
| De façon sévère | durement, lourdement, rigoureusement, sévèrement | 0 | Ces collocatifs adverbiaux n'ont pas d'équivalents directs en arabe. On peut employer le verbe 2 تأدب [<i>'addaba</i>] soit avec une expression libre composée de la préposition ب [<i>b</i>] + le <i>masdar</i> شِدَّة/صِرَامَة/قَسْوَة [<i>šidda/šarāma/qaswa</i>], soit avec une collocation composée du <i>masdar</i> 2 تأديب 2 [<i>ta'dīban</i>] (mis à l'accusatif indéfini parce qu'il fonctionne en tant qu'objet interne) + l'un des adjectifs suivants (mis également à l'accusatif indéfini) : قَسِيًّا / شَدِيدًا / صَارِمًا / سَارِمًا [<i>šadīdan/šarīman/qāsīyan</i>]. Pour plus d'informations, consulter le dictionnaire unilingue arabe. | أدب المعلم على تصرفه السيئ تأديبًا صارمًا. [<i>'addaba-hu al-mu'allimu 'alā tašarrufi-hi as-sayyi' ta'dīban šarīman</i>]. Lit. 'L'instuteur l'a châtié d'un châtiement sévère pour son mauvais comportement.' |

| | | | | |
|--------------------|-----------------|---|---|---|
| De façon justifiée | bien, justement | 0 | <p>Ces collocatifs adverbiaux n'ont pas d'équivalents directs en arabe. On peut employer le verbe 2 أَدَبَ avec une collocation composée de l'adjectif comparatif أحسن [aḥsana] (mis à l'accusatif) + le masdar 2 تَدَبَّى [ta'dib].</p> <p>On peut aussi employer le verbe 2 أَدَبَ avec une expression libre comme, par exemple, « كما يجب » [kamā yajīb] qui signifie 'comme il se doit'.</p> | <p>أَدَبَ المعلم أحسن تَدَبَّى. [aḍḍaba-hu al-muʿallimu aḥsana ta'dib]. Lit. "l'instituteur l'a châtié d'un meilleur châtiement."</p> |
|--------------------|-----------------|---|---|---|

Exemples de traduction

1- On ne doit pas châtier les enfants avant qu'ils puissent connaître leurs fautes.
[min al-mafrūḡ 'an lá nu'tadib al-'aṭfál qabla 'an nubayyin lahum aḥtá'ahum].
من المفروض أن لا تؤدب الأطفال قبل أن نبين لهم أخطأهم.

2- L'inceste est universellement prohibé et même sévèrement châtié dans toutes les sociétés.
[ǧiṣyānu al-maḥāhīm maḥzūr wa jarīma yu'āqib 'alayha bi-šidda fi jamā' al-mujtama'āt].
عشيان المخارم محظور و جريمة يُعاقب عليها بشدة في جميع المجتمعات.

Remerciements

Je tiens à remercier Gaétane Dostie et Alain Polguère d'avoir lu et commenté une première version de cet article.

Bibliographie

- Alonso Ramos (M.), 2001 : « Construction d'une base de données des collocations bilingue français-espagnol », dans *Langages*, 143, p. 5-27.
- Baker (C. F.) *et al.*, 1988 : « The Berkeley FrameNet Project », dans *Proceedings of COLING-ACL'98*, Montréal, p. 86-90.
- Bouchaddakh (S.), 2010, en préparation : *Application des principes de la lexicologie explicative et combinatoire à la construction d'un dictionnaire bilingue informatisé français-arabe*, thèse de doctorat dirigée par G. Dostie et A. Polguère, Université de Sherbrooke.
- Bouchaddakh (S.), 2008 : « La définition dans les dictionnaires bilingues : problèmes de polysémie et d'équivalence interlangues », dans *Proceedings of the XIII EURALEX International Congress: EURALEX'2008*, Barcelone, p. 807-812.
- Fellbaum (C.), 1998 : *WordNet: An Electronic Lexical Database*, Cambridge (MA), MIT Press.
- Fillmore (C.) *et al.*, 2001 : « Building a Large Lexical Databank Which Provides Deep Semantics », dans *Proceedings of the Pacific Asian Conference on Language, Information and Computation*, Hong Kong, p. 3-26.
- Fontenelle (T.), 1997 : *Turning a Bilingual Dictionary into a Lexical Semantic Database*, Tübingen, Max Niemeyer Verlag.
- Hausmann (F. J.), 1979 : « Un dictionnaire des collocations est-il possible? », dans *Travaux de linguistique et de littérature*, 17 (1), p. 197-195.
- Mel'čuk (I. A.) *et al.*, 1995 : *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, Duculot.
- Mel'čuk (I. A.), 1996 : « Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon », dans Wanner (L.), éd., *Lexical Functions in The Lexicography and Naturel Language Processings*, Amsterdam/Philadelphia, Benjamins, p. 37-102.
- Mel'čuk (I. A.) et Polguère (A.), 2007 : *Lexique actif du français: L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*, coll. « Champs linguistiques », Bruxelles, De Boeck & Larcier.
- Selva (T.) *et al.*, 2002 : « Le DAFLES, un nouveau dictionnaire pour apprenants du français », dans *Proceedings of the Tenth EURALEX International Congress: EURALEX'2002, Copenhagen, Vol.1, p. 199-208*.
- Polguère (A.), 2000 : « Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French », dans *Proceedings of the Ninth EURALEX International Congress: EURALEX'2000*, Stuttgart, p. 517-527.
- Polguère (A.), 2003 : « Étiquetage sémantique des lexies dans la base de données DiCo », dans *Traitement Automatique des Langues (T.A.L.)*, vol. 44, n° 2, p. 39-68.

Verlinde (S.) et al., 2005 : « Dictionnaires électroniques et environnement d'apprentissage du lexique », dans *Revue de linguistique appliquée*, vol. 10, n° 2, p. 19-30.

Dictionnaires consultés

Al-Baalbaki (M.), 1982 : *Al-Mawrid a Modern English-Arabic Dictionary*, Beirut, Dar al-ilm lil-malayin.

Al-Najjari Bek (M.), 1970 : *Dictionnaire français-arabe*, Beyrouth, Librairie du Liban.

Blachère (R.), Chouemi (M.), et Denizeau (C.), 1967 : *Dictionnaire arabe-français-anglais: langue classique et moderne*, Paris, G.-P. Maisonneuve et Larose.

Idris (S.), 2001 : *Al-Manhal dictionnaire français-arabe*, Beirut, Dar al-Adab.

Kasimirski, 1860 : *Dictionnaire arabe-français*, Beyrouth, Librairie du Liban.

Reig (D.), 1983 : *As-Sabil Dictionnaire arabe-français, français-arabe*, Paris, Larousse.

Reig (D.), 1999 : *Dictionnaire arabe-français, français-arabe*, Paris, Larousse.

Torkia (E.), 1952 : *Nouveau dictionnaire français-arabe*, Caire, Bureau de la Publicité.

Miloud Taïfi, Patrice Pognan,
Labo-Ling (U. Fès) et LALIC (U. Paris-Sorbonne et INALCO),
Université Sidi Mohamed ben Abdellah et INALCO,
Fès et Paris,
Maroc et France,
miltafi@yahoo.fr et mcertal@wanadoo.fr.

Le Dictionnaire raisonné berbère-français. ***Parlers du Maroc et ses traitements informatiques***

Résumé

Le Dictionnaire raisonné berbère-français. Parlers du Maroc, basé sur une écriture phonologique et grammaticale, a pour but premier de reconstituer le système de la langue berbère. Bien structuré, il permet des traitements informatiques grâce auxquels un index des correspondances mots berbères/racines berbères et un lexique significations françaises/racines berbères ont pu être adjoints au dictionnaire. Ce dictionnaire transformé en corpus permet la construction d'une base de données dont le but principal sera de préparer la masse lexicale nécessaire à l'élaboration d'un dictionnaire français-berbère.

Mots-clés: lexicographie; bilinguisme berbère-français; dictionnaire; système de la langue berbère; écriture phonologique; traitements informatiques du dictionnaire.

1 Présentation du dictionnaire

Ce dictionnaire est une version corrigée, augmentée et remaniée de l'ouvrage intitulé *Dictionnaire tamazight-français (parlers du Maroc central)* publié en 1992. Le but visé étant de reconstituer le système de la langue berbère, cette nouvelle version est suffisamment différente de la première, dans plusieurs de ses aspects, pour en proposer un autre titre qui est désormais : *Dictionnaire raisonné berbère-français. Parlers du Maroc*. Cela veut dire que la masse lexicale consignée dans cet ouvrage n'est plus, comme dans le précédent, confinée au dialecte tamazight qui regroupe les parlers pratiqués dans le Maroc central, mais comporte aussi, comme il est expliqué ci-après, des formes de mot et/ou des significations nouvelles attestées dans d'autres zones géolinguistiques berbérophones du Maroc.

1.1 Correction

La correction a consisté à redresser les fautes, les erreurs et les maladroites aussi bien dans la partie berbère du dictionnaire que dans sa partie française. La saisie de l'ouvrage a permis de revoir avec plus de détail et d'acuité toutes les scories que comporte l'ancienne version. Les supports informatiques et leur manipulation ont

été d'un grand secours. Il est cependant évident, malgré toute l'attention portée à la réécriture de l'ouvrage, que le lecteur peut trouver encore quelques coquilles ou quelques oublis.

1.2 Augmentation

Deux sources essentielles nous ont permis de procéder à l'augmentation de la masse lexicale répertoriée dans cet ouvrage.

Depuis 1992, la lexicographie berbère marocaine a connu un essor remarquable de par la réalisation de divers travaux sur le lexique, travaux académiques en majorité; ce qui nous a permis de renouveler et d'augmenter le dictionnaire précédent (Oussikoum 1995, Azdoud 1997, Boumalk et Bounfour 2001, Serhoual 2002, Rahho 2005). Nous avons ainsi, en puisant, avec prudence et circonspection, dans ces travaux, enrichi le dictionnaire d'autres mots, d'autres expressions et d'autres acceptions, en élargissant l'investigation à d'autres parlers du Maroc, appartenant aux dialectes dits « tarifit » ou « tachelhit », en nous fondant essentiellement sur les critères de production et/ou de réception et en considérant la langue berbère du Maroc dans sa globalité, espérant, de par cette orientation, participer à l'effort collectif de la propagation du berbère auprès de tous ses locuteurs. On remarquera ainsi que nous avons renoncé à indiquer l'appartenance des formes à tel ou à tel parler ou dialecte, mettant en évidence, de cette façon, le système de la langue en elle-même et non pas les différentes et diverses performances de locuteurs. L'option de transcription adoptée dans ce dictionnaire, comme il est expliqué ci-après, renforce encore une telle orientation;

La seconde source concerne les corpus de littérature orale, dans toute la diversité de celle-ci : chants, poésie, proverbes, devinette, contes... Textes authentiques, les formes littéraires sont des garanties d'attestation. Nous avons exploité ainsi des documents publiés et plusieurs autres corpus collectés par des étudiants dans le cadre de leurs travaux académiques en thèse (Amrani 2007, Kich 2007, Jarmouni 2009). Une source inestimable que la littérature, car elle conserve des mots, des expressions et des acceptions que l'usage quotidien du berbère n'actualise jamais ou du moins qu'elle n'actualise que rarement ! L'exploitation de la littérature nous a permis aussi de diversifier les exemples du dictionnaire en y insérant d'autres chants, proverbes et devinettes.

1.3 Remaniement

Nous avons adopté de façon plus systématique une écriture phonologique et grammaticale. La première option consiste à reconstituer, quand cela est possible, les éléments constitutifs de la racine qui subissent, lors des réalisations phonétiques, des changements et des altérations dans les formes de mot ou à la frontière des constituants au niveau des séquences syntagmatiques. Les changements phonétiques ne sont pas nombreux, ils portent surtout sur des réalisations circonscrites dans l'aire linguistique du berbère. Quelques exemples suffiront à montrer de quoi il s'agit :

- [1] le passage de *k* à *š*: *akal* > *ašal* « terre, sol » ;
- [2] le passage de *g* à *ž*: *igenna* > *iženna* « ciel » ;
- [3] le passage de *l* à *ž*: *alim* > *ažim* ou bien à *r*: *alim* > *arim*.

Ces changements n'altèrent en rien la structure des racines. Par contre, la vocalisation des semi-consonnes a un effet corrosif. En effet *y* est souvent, dans certaines formes de mot, réalisé en la voyelle *i*. Il en est ainsi par exemple du verbe *asy* « prendre », dont la racine constitutive est bilitère *SY*, qui voit sa deuxième radicale actualisée en *i* dans certaines formes de conjugaison :

- [4] *usix* au lieu de *useyx* « j'ai pris » ;
- [5] *tusim* au lieu de *tuseym* « vous avez pris ».

Ce qui réduit la forme verbale à une racine monolittère. Il en est de même pour la semi-consonne *w*, réalisée en la voyelle *u* dans quelques contextes phoniques. Ainsi, le verbe *arw* « enfanter, accoucher » est trilitère, mais dans certaines de ses formes conjuguées *w* est réalisé *u* :

- [6] *turud* « tu as accouché » au lieu de *turewd*

Ce qui réduit là aussi la racine bilitère à une monolittère.

La seconde option, qui constitue sans doute une innovation, consiste à rendre transparents et visibles dans l'écriture tous les éléments de la langue, lexicaux ou grammaticaux, constitutifs des énoncés. Une telle écriture permet ainsi l'identification des objets linguistiques tels qu'ils se présentent dans le système de la langue, indépendamment des performances aussi variées des locuteurs berbérophones.

1.4 Reconstitution du système de la langue

Le principe fondamental qui préside à la méthodologie appliquée dans ce dictionnaire est la reconstitution du système de la langue berbère, telle qu'elle se présente dans le domaine marocain. On est en effet très loin de l'époque où l'on affirmait que le berbère est constitué d'une « poussière » de parlers, chacun confiné dans une zone géographique limitée, réduite parfois à un mouchoir de poche. Plusieurs facteurs sociaux ont depuis contribué à l'ouverture des parlers et à leurs contacts avec les autres : le mouvement des populations au Maroc, une sédentarisation accélérée, les médias de masse (la radio notamment), le tissu associatif défendant la cause berbère et, plus récemment, l'insertion du berbère dans le système éducatif ont réduit sensiblement l'étanchéité entre les parlers et les dialectes. Ajoutons aussi que les recherches académiques ont permis une connaissance plus approfondie de plusieurs parlers en révélant leurs particularismes. La reconstitution du système impose de ce fait la méthodologie appliquée dans la confection de ce dictionnaire. Trois options en sont les plus saillantes :

- notation des formes de base selon leur phonie initiale, en procédant cependant à des renvois, quand cela est nécessaire, à des réalisations réelles particulières ;
- dissimilation des complexes phonétiques au niveau de toutes les séquences syntagmatiques dans lesquelles ils apparaissent ;
- traitement des variétés lexicales dans le cadre de la synonymie ou parasynonymie en considérant que toutes les formes de mot rapprochées de par leurs affinités de sens, appartiennent au lexique de la langue berbère, indépendamment de leur actualisation dans tel ou tel parler.

C'est ce qui justifie le qualificatif « raisonné » inclus dans le titre de cet ouvrage.

1.5 Comparaisons

1.5.1 Comparaison avec le kabyle

Il nous a paru utile de maintenir la comparaison des données de ce dictionnaire avec le kabyle en nous référant exclusivement à Dallet (1982). La comparaison pourrait servir aux études dialectologiques. Elle nous montre que les différents dialectes se partagent un grand nombre de racines. Mais les mots formés d'une même racine ne recouvrent pas toujours les mêmes sens. Le rapprochement n'a été noté que dans des cas où les parlers berbères du Maroc et le kabyle présentent au moins un sens commun, un invariant de sens pour une même racine.

1.5.2 Comparaison avec l'arabe

Le berbère a admis beaucoup de vocables étrangers. On y trouve des mots latins, turcs, français, espagnols..., mais ce sont surtout les emprunts faits à l'arabe qui constituent la plus grande partie des apports étrangers. On sait par ailleurs que l'arabe et le berbère appartiennent à la même famille de langues : le chamito-sémitique. Les deux systèmes contiennent donc nécessairement un fonds lexical commun. Dans l'état actuel, le berbère et l'arabe dialectal marocain sont en contact étroit. Il y a donc inévitablement emprunt de part et d'autre, et il n'est pas toujours aisé de statuer sur la provenance de certaines racines.

Il est évident cependant que des sous-systèmes lexicaux relatifs à des domaines particuliers : religieux, sociopolitique..., sont empruntés à l'arabe. Mais l'examen de l'ensemble lexical montre que les racines communes aux deux systèmes ne sont pas toutes exclusivement arabes. Les rapprochements notés dans ce dictionnaire n'indiquent donc pas l'origine des racines berbères mais indiquent par contre qu'il y a, en synchronie, simple similitude entre le berbère et l'arabe.

1.5.3 Référence à d'autres langues

Nous indiquons aussi, de façon sporadique et quand cela nous semble plausible, l'origine des emprunts faits à d'autres langues, notamment au français, et l'espagnol, au latin et au turc.

1.6 Traduction en français

La traduction d'une langue en une autre n'est pas une opération facile, et ce travail n'échappe pas aux problèmes auxquels est confrontée toute étude de lexicographie différentielle. La tâche a été d'autant plus malaisée que le berbère et le français sont deux langues qui appartiennent à des familles distinctes et représentent des cultures foncièrement différentes.

Les difficultés de la traduction relèvent en effet de deux sortes de causes :

- des causes d'ordre linguistique : les mots ne sont pas équivalents et ne recouvrent pas toujours les mêmes acceptions ;
- des causes d'ordre culturel : les langues expriment différemment les réalités environnantes.

Les exemples retenus servent à illustrer chaque sens et montrent aussi comment le berbère organise les éléments linguistiques au niveau de la chaîne. La traduction littérale est donnée dans des cas où il y a ambiguïté sémantique ou pour mettre en relief un phénomène de syntaxe. Les expressions, locutions, proverbes et pièces poétiques sont d'abord, pour la plupart, traduits littéralement pour montrer la différence entre leur sens littéral et leur signification globale.

1.7 Classification par racines et organisation des articles sous une même racine

Les racines dégagées sont classées suivant l'ordre alphabétique du français adapté aux phonèmes particuliers du berbère. Beaucoup de racines sont homonymes, c'est-à-dire composées des mêmes consonnes. L'homonymie concerne surtout les monolitères, les bilitères et plus rarement les trilitères. L'ordre de classification des racines homonymes est le suivant : ont été notées d'abord celles qui fournissent les outils grammaticaux : pronoms, particules, conjonctions..., ensuite les racines verbo-nominales et, en dernier lieu, les racines qui sont exclusivement nominales.

Chaque racine dégagée est indiquée en lettres capitales. Elle constitue l'entrée-vedette d'un ou de plusieurs articles. En face de la racine, à droite sur la même ligne, sont indiqués soit l'origine, quand elle est bien établie, soit les rapprochements avec d'autres langues, notamment, ce qui est plus fréquent, avec l'arabe et/ou le kabyle.

Chaque article est introduit au début de la ligne par le signe ◆ ; sont notées immédiatement après, à l'aide de la barre oblique/les variantes phonétiques ou morphologiques.

Le signe ◆ introduit les sens. Les exemples viennent ensuite, précédés du signe ●. La traduction est séparée de l'exemple par une simple virgule. La traduction littérale ou une note explicative sont toujours mises entre parenthèses.

1.8 Contenu

Les traitements informatisés (dans un environnement de programmation en Python) réalisés sur le dictionnaire permettent d'obtenir un certain nombre d'indications chiffrées. Mis « à plat », ce dictionnaire représente 2 700 000 caractères, plus de 40 000 lignes (enregistrements informatiques), près de 7 200 racines (ce qui constitue une augmentation d'environ 50 % du nombre de racines par rapport au dictionnaire de 1992) et près de 18 000 articles, dont 5 000 concernent des verbes simples ou dérivés. C'est aussi un ensemble de 13 500 exemples et locutions berbères authentiques. Ce nouveau dictionnaire couvre ainsi pratiquement le lexique commun de tous les parlers berbères du Maroc.

2 Les traitements informatiques

- Ces traitements sont de deux types :
- ceux qui complètent le dictionnaire par l'ajout d'index ;
- ceux qui utilisent le dictionnaire pour aborder des développements nouveaux.

2.1 Création d'index

Le dictionnaire étant classé par racine suivant la tradition des études sémitiques, il a semblé opportun de proposer au lecteur la possibilité de circuler plus aisément au sein du dictionnaire par l'ajout d'un index « mot berbère/racine berbère », la détermination de la racine à partir du mot n'étant pas toujours aisée. Nous avons également souhaité fournir un accès à ce dictionnaire par le français grâce à un index « français/racine berbère », ce qui permet d'attendre la réalisation du « Taïfi » français-berbère.

Le Dictionnaire raisonné berbère-français est bien structuré, ce qui facilite grandement les traitements informatiques. Les différents fichiers Word ont été transformés en fichiers bruts codés en Unicode UTF-8. Ils ont été fusionnés en un seul fichier qui sert de base à tous les traitements ultérieurs. L'exemple d'un article de dictionnaire donné ci-dessous (figure 1) permettra de comprendre les traitements réalisés.

L'extraction des articles permet de créer le premier index « mot berbère/racine berbère » alors que l'extraction et le traitement des éléments en gris fournit les entrées de l'index « français/racine berbère ».

Classification des racines et organisation des articles

BDR

◆ budr

budr, tbudur, ur-budr ► se maquiller, se farder ; se faire beau ; soigner son apparence. (v. aussi : **helleg, hlq ; herges, hrs**). ● la **tbudur hma** ad **teddu ger tmegra**, elle se maquille, elle se fait belle pour aller assister à la fête du mariage.

◆ S — sbudr

sbudr, tesbudur, ur-sbudr ► maquiller, farder, mettre du fard (s'emploie avec la prép. i, à). ● **tesbudr i yilli-s**, elle a maquillé sa fille.

◆ abudr (u)

ibudren ► maquillage (résultat, état). ● **ur as igudi ubudr**, elle ne s'est pas bien maquillée.

◆ abudir (u)

ibudurn ► fard, produit qui sert à se maquiller (crème, rouge à lèvres ...)

◆ asbudr (u)

isbudren ► action de maquiller, de farder ; maquillage (action).

BDR

◆ abadder (u)

Figure 1

2.1.1 L'index « mot berbère/racine berbère »

Le passage du fichier dictionnaire tel qu'il est ci-dessus à l'index, et dont un extrait est présenté ci-dessous (figure 2), est décomposé en 15 tâches exécutées par de petits programmes. Ils concernent essentiellement le traitement de la ponctuation et l'analyse des formes verbales dérivées.

Lexique berbère

| | | | |
|----------------|--------------|---------------|-----------------|
| izig: ZX1 | žawr: ŽWR2 | žewwa: ŽW1 | žžel: ŽL2 |
| izikr: ZKR3 | žawr: ŽWR3 | žewweq: ŽWQ1 | žželbana: ŽLBN1 |
| izim: ZM2 | žber: ŽBR1 | žeyyef: ŽYF1 | žželf: ŽLF1 |
| izimmer: ZMR4 | žber: ŽBR2 | žeyyer: ŽYR1 | žželłita: ŽLł1 |
| izimr: ZMR4 | ždeb: ŽDB1 | žecžec: ŽE1 | žženžar: ŽNžR1 |
| izinfer: ZNFR2 | ždeb: ŽDB2 | žecleq: ŽELQ1 | žžent: ŽN4 |
| izl: ZL1 | žebben: ŽBN1 | žecłet: ŽEł1 | žžerda: ŽRD2 |
| izli: ZL8 | žebber: ŽBR3 | žfel: ŽFL1 | žžerđa: ŽRD2 |
| izm: ZM2 | žedder: ŽDR1 | žgem: ŽGM1 | žžernan: ŽRN1 |
| izrezzi: ZRž1 | žeffef: ŽF3 | žhaž: ŽHž1 | žžert: ŽR3 |
| izri: ZR11 | žeffen: ŽFN1 | žhed: ŽHD1 | žžerł: ŽR12 |

Figure 2

2.1.2 L'index « français/racine berbère »

Le lexique français vers la racine berbère est réalisé en 11 programmes et 2 étapes manuelles. Le commentaire de la démarche adoptée s'appuiera sur l'exemple donné en gris clair dans la première illustration :

[7] ► se maquiller, se farder ; se faire beau ; soigner son apparence. (v. aussi : *helleq*, *hlq* ; *hergeş*, *hrğş*).

La première étape ôte le triangle qui annonce une signification ainsi que les renvois qui seront traités par ailleurs pour obtenir automatiquement les synonymes. La seconde étape extrait les sens principaux limités par le point-virgule, séparateur univoque au sein du dictionnaire. On obtient ainsi :

[8] se maquiller, se farder
se faire beau
soigner son apparence

L'étape qui suit est la première étape manuelle au cours de laquelle est effectué le choix des parenthèses à conserver (cf. exemple figure 3).

| Lexique des significations françaises | | |
|--|-----|---|
| <u>belle-mère</u> | B5 | |
| <u>marâtre</u> (seconde femme du père) | B5 | |
| <u>plante</u> du genre de la fêrùle, très recommandée pour le bétail. | B6 | |
| <u>porte</u> , portail, planche qui sert de porte | B7 | |
| <u>entrée</u> , ouverture | B7 | |
| <u>fig.</u> chance, bonheur, bon augure | B7 | |
| <u>petite porte</u> , ouverture par laquelle on fait entrer les moutons dans l'enclos. | B7 | |
| <u>portier</u> , factionnaire à l'entrée d'un ksar. | B7 | |
| <u>moustique</u> , cousin (insecte) | B8 | |
| <u>mamelle</u> , sein (généralement gros sein) | B9 | |
| <u>petite mamelle</u> , petit sein, sein de jeune fille. | B9 | |
| <u>sein</u> | B9 | traitement manuel: choix des parenthèses à conserver |
| <u>dindons</u> | B10 | |
| <u>dindon</u> | B10 | |
| <u>dinde</u> , femelle du dindon. | B10 | |
| | | |

Figure 3

On suppose que le lecteur francophone connaît le mot « marâtre » qui n'est pas ambigu et peut être recherché dans un dictionnaire du français. La parenthèse « généralement gros sein » n'est pas conservée. Par contre, la mention « insecte » liée à « cousin » est primordiale pour la recherche dans le dictionnaire.

Cela donne, à terme, des résultats de ce type :

| | |
|------------------|----|
| [9] belle-mère | B5 |
| marâtre | B5 |
| férule (plante) | B6 |
| moustique | B8 |
| cousin (insecte) | B8 |
| mamelle | B9 |
| sein | B9 |

Cette étape ne peut pas être conduite automatiquement, car elle concerne des choix sémantiques pour lesquels il faudrait posséder un traitement automatique général de la sémantique du français qui, à notre connaissance, n'existe pas (encore).

Après divers aménagements, en particulier des mentions du type « mm. sens que préc. » ou « dim. du préc. », le programme coupe les sens secondaires sur la base de la virgule (qui n'est pas un séparateur fiable) et propose un fichier des significations françaises face aux racines berbères (numérotées) correspondantes. Ce fichier est dépouillé manuellement, puis après un tri la chaîne de programmes reprend : les enregistrements identiques sont réduits à une seule occurrence et les racines multiples pour une signification française sont regroupées. Nous donnons ci-après un exemple de résultats (figure 4) :

| significations françaises - racines berbères | |
|--|---|
| <u>aède</u> : DYZ1 | <u>alors</u> : D7, DY1 |
| <u>affaibli</u> : DHŠ1 | <u>amabilité</u> : DRF3 |
| <u>affaiblir</u> : DR2 | <u>amant</u> (prendre pour ~): DKL1 |
| <u>affaire</u> (se tirer d'~): DBR1 | <u>ambitionner</u> : DMÉ1 |
| <u>affection cutanée</u> : DD3 | <u>amende</u> : DÉR1 |
| <u>afin que</u> : BŠ1 | <u>amitié</u> (se lier d'~ avec qn): DKL1 |
| <u>agaçant</u> : BSL1 | <u>amonceler</u> : DKN1 |
| <u>agent de police</u> : BLS3 | <u>amour</u> : BD7, YR2 |
| <u>agglomération rurale</u> : DŠR2 | <u>ampoule</u> : BQL1 |
| <u>agir injustement envers qn.</u> : DLM1 | <u>ancien</u> : BLY1 |
| <u>agir sans réfléchir</u> : DRM1 | <u>anémone</u> : BN&MN1 |
| <u>agissement irréflechi</u> : DRM1 | <u>anesthésie</u> : BNŽ1 |
| <u>agitation</u> : DGY2, DRW1 | <u>animal domestique</u> : BHM1 |

Figure 4

Les traitements manuels effectués apparaissent bien dans ces exemples. À partir d'expressions telles que :

- [10] prendre pour amant
 se tirer d'affaire
 se lier d'amitié avec quelqu'un

nous produisons automatiquement les entrées « prendre », « se tirer » et « se lier », mais aussi par intervention manuelle les entrées « amant », « affaire », « tirer (se) », « amitié » et « lier (se) » afin d'assurer le meilleur accès à l'étude du vocabulaire afférent à une racine.

2.2 Développements nouveaux

2.2.1 « L'exemplier Taïfi »

L'ensemble des exemples du *Dictionnaire raisonné berbère-français* a été extrait, soit aux environs de 13 500 exemples. Nous appelons ce fichier « l'exemplier Taïfi ». Il se présente de la manière suivante :

- iferred g isekwla alliy qqurn, il a négligé les arbres (fruitiers), et ils ont séché (ils sont devenus secs).
- ituferred g umudin alliy immut, le malade a tellement manqué de soins qu'il en est mort.
- la iferfid xef teflew t g tillas, il cherche la porte à tâtons dans le noir.
- ar iferfid illi-s n eemmi-s ddaw n yiherbel, il caressait discrètement sa cousine sous la couverture.
- ifreg s uzeggwar i yišerwan, il a construit un enclos pour agneaux avec du jujubier sauvage.
- ifreg i waman, il a mis une digue pour dévier le cours d'eau.
- ifreg-as eemmi-s alliy yusy ixf-nes, son oncle paternel l'a pris sous sa tutelle jusqu'à sa majorité.
- ad ax ifreg rebbi seg wallen, que Dieu nous protège du mauvais œil (des yeux).
- freg i wussan-nek (protège tes jours), préserve ton âme du châtime (conseil donné au calomniateur).

Figure 5

Un programme de découpage de cet exemplier en expression berbère- (traduction littérale)-traduction est révisé de manière systématique et sera réintroduit dans le programme de fabrication de la base de données à partir du dictionnaire. Ce programme s'appuie sur l'analyse de la ponctuation et la présence de caractères propres au berbère d'une part, et propres au français d'autre part, ainsi que sur quelques mots outils particulièrement fréquents. Le résultat est un fichier brut bilingue berbère-français. À partir de celui-ci un second fichier sera produit : celui, monolingue, des exemples berbères.

Ces deux fichiers revêtent une grande importance. Ils permettront, en effet, de faire de nombreuses recherches linguistiques. Ils serviront également de corpus pour l'élaboration de programmes d'analyse automatique du berbère concernant le lexique, la morphologie et la syntaxe. Actuellement, un programme d'extraction

automatique du syntagme nominal prépositionnel est en cours de réalisation. Il sera testé au plus vite sur le fichier monolingue.

2.2.2 Constitution automatique d'une base de données

Elle est réalisée en deux temps. La première série de programmes a pour but de mettre en évidence la structure implicite du dictionnaire, ce qui donne le résultat suivant (figure 6) :

| | |
|-----|--|
| BDR | <ul style="list-style-type: none"> ◆ budr, budr, tbudur, ur-budr <ul style="list-style-type: none"> ▶ se maquiller, se farder ; se faire beau ; soigner son apparence. (v. aussi : helleq, h1q ; herges, hrgs). • la tbudur hma ad teddu ger tmegra, elle se maquille, elle se fait belle pour aller assister à la fête du mariage. ◆ S — sbudr, sbudr, tesbudur, ur-sbudr <ul style="list-style-type: none"> ▶ maquiller, farder, mettre du fard (s'emploie avec la prép. i, à). • tesbudr i yilli-s, elle a maquillé sa fille. ◆ abudr (u), ibudren <ul style="list-style-type: none"> ▶ maquillage (résultat, état). • ur as igudi ubudr, elle ne s'est pas bien maquillée. ◆ abudir (u), ibudum <ul style="list-style-type: none"> ▶ fard, produit qui sert à se maquiller (crème, rouge à lèvres ...) ◆ asbudr (u), isbudren <ul style="list-style-type: none"> ▶ action de maquiller, de farder ; maquillage (action). |
| BDR | <ul style="list-style-type: none"> ◆ abadder (u), ibaddem <ul style="list-style-type: none"> ▶ pied d'une pente, talus, petite colline ; mur, mur d'une clôture. ◆ asebdir (u), isebdim <ul style="list-style-type: none"> ▶ plaine présentant de légers mouvements de terrain. |

Figure 6

Cette forme du fichier du dictionnaire sert de point de départ à la deuxième série de programmes qui vont apporter des informations supplémentaires, par exemple calcul des formes d'annexion du substantif (quel que soit le genre et le nombre) et de l'adjectif, analyse des exemples en exemple berbère, traduction littérale éventuelle et traduction...

Le premier programme traite de la morphologie de la manière suivante :

Il reconnaît les verbes simples et fabrique le grammatème « VN (dérivation verbo-nominale) – V (verbe) – (1.1) expression du verbe simple ». La génération automatique de l'échelle numérique (« 1.1 », « 1.2 », « 1.3 », « 2 », où le « 1 » renvoie à une dérivation verbo-nominale et le « 2 » à une dérivation nominale) permettra la sortie à partir de la base de données d'un dictionnaire par racines qui classe les différents mots sous la racine correspondante suivant les valeurs de cette échelle attribuées aux mots.

Il produit aussi :

- les verbes dérivés: « VN – V – 1.2 » ;
- les substantifs verbaux: « VN – SUBST – 1.3 » ;
- « nomen » (substantifs et adjectifs) avec une dérivation nominale, soit :
- « N – SUBST – 2 » ou « N – ADJ – 2 ».

et par la fin pour le français, le programme reconstitue, en très grande partie, les suites berbères et françaises.

Etape 3: traitement des exemples

| | | | | | | | | |
|----------------------|------|----------|---|------|--|---------------------|---------|------------------------------|
| BDL | BDL1 | ar., kb. | | | | | | |
| | ◆ 1 | VN | V | 1.1. | beddel | beddel | tbeddal | ur-beddil |
| | | | | | ▶ changer, varier (trans. et intrans.), transformer, se transformer, modifier, être modifié. | | | |
| | | | | | ▶ remplacer, être remplacé, échanger, substituer. | | | |
| | | | | | • ibeddel ṛay | | | la décision a été annulée |
| | | | | | • ibeddel wudm-nes | son visage a chargé | | il a maigri / il est devenu |
| subitement pâle | | | | | | | | |
| ne le reconnais plus | | | | | • ibeddel ġif-i | il a changé sur moi | | il a tellement changé que je |
| | | | | | • ibeddel awal-nes | | | il a contrefait sa voix |
| un mulet | | | | | • ibeddel aġyul s userdun | | | il a échangé un âne contre |

Figure 8

Conclusion

L'existence du dictionnaire raisonné berbère – français et les traitements informatiques qui lui sont appliqués permettent d'envisager plusieurs directions de travaux et de recherches.

L'élaboration d'une base de données à partir du dictionnaire, en dehors des questions de consultation, a pour but premier de permettre de construire progressivement un dictionnaire français – berbère correspondant au dictionnaire actuel. Cette construction est réalisée par une procédure de retournement automatique de la masse lexicale intégrée dans une nouvelle base de données orientée du français vers la langue étrangère (procédure déjà testée sur un lexique slovaque-français). Cela n'économise en rien les processus de conception du dictionnaire, mais en accélère considérablement la réalisation.

L'expérience acquise dans la génération automatique des formes de l'état d'annexion lors de la construction de la base de données sert déjà dans la réalisation d'un programme d'analyse du syntagme nominal prépositionnel.

La présence du *Dictionnaire raisonné berbère-français* est l'élément déterminant qui a permis d'envisager le lancement de travaux d'analyse automatique du berbère.

Les différents modules qui en résulteront seront testés sur l'exemplier et des textes, en particulier saisis optiquement.

La réalisation d'un gestionnaire de fichiers bruts (en Unicode UTF-8), Brutus, est désormais à l'ordre du jour dans le but de faciliter la consultation de textes berbères et de l'exemplier. Des programmes de manipulation des textes seront progressivement implémentés et les modules d'analyse du berbère seront introduits dans ce gestionnaire au fur et à mesure de leur réalisation dans le but de faciliter les recherches linguistiques sur le berbère.

Bibliographie

- Amrani (F) 2007 : *Le lexique berbère du corps humain (Maroc central), Approche sémantique et symbolique*. Thèse de doctorat d'État. Faculté des Lettres et des Sciences Humaines, Dhar El Mehraz, Fès.
- Azdoud (D) 1997 : *Lexique commun des Ait Hadiddou du Haut Atlas*, Thèse de doctorat d'État, Faculté des Lettres et des Sciences humaines. El Jadida.
- Boumalk (A.) et Bounfour (A.) 2001 : *Vocabulaire usuel de tachelhit, tachelhit-français*, éd. Centre Tarik Ibn Zyad, Rabat.
- Jarmouni (M.-H.) 2009 : *Anthologie analytique de la poésie berbère (tamazight) du Moyen Atlas*, Thèse de doctorat, Faculté des Lettres et des Sciences Humaines, Dhar El Mehraz, Fès.
- Kich (A.) 2007 : *De la poésie orale berbère (tamazight): typologie et mutations*, Thèse de doctorat d'État, Faculté des Lettres et des Sciences Humaines, Marrakech.
- Oussikoum (B) 1995 : *Dictionnaire tamazight-français (parler des Ayt Wirra)*, Thèse de doctorat d'État, Faculté des Lettres et des Sciences Humaines, Beni-Mellal.
- Rahho (R) 2005 : *Dictionnaire berbère-français. Parler des Beni-Iznassen*, Thèse de doctorat, Faculté des Lettres et des Sciences Humaines, Dhar El Mehraz, Fès.
- Serhoual (M) 2002 : *Lexicologie, lexicographie et sémantique berbères : 1) Lexicologie amazighe 2) Dictionnaire tarifit-français*, Thèse de doctorat d'État, Faculté des Lettres et des Sciences Humaines, Tétouan.

Jorge Baptista,
Universidade do Algarve — FCHS,
Faculdade de Ciências Humanas e Sociais,
Faro,
Portugal,
jbaptis@ualg.pt.

Dolors Català,
Universitat Autònoma de Barcelona,
Departament de Filologia Francesa y Románica,
Barcelona,
Espanne,
Dolors. Catala@uab.cat.

Adverbes focalisateurs et analyse syntaxique automatique de groupes nominaux

Résumé

À partir d'une classification des adverbes en *-ment* du français, nous nous sommes proposés de donner une description équivalente des adverbes du portugais (PT) et de l'espagnol (ES) et de les comparer. Il sera question ici des adverbes focalisateurs (notés MF), dont nous envisageons d'intégrer la description linguistique dans des analyseurs syntaxiques, dans le cadre du traitement automatique des langues (TAL) en vue de l'amélioration de la description syntaxique profonde (*deep parsing*).

Mots-clés : adverbe focalisateur ; traitement automatique des langues (TAL) ; analyseur syntaxique (parser) ; syntaxe comparative ; portugais ; espagnol.

Introduction

Les adverbes focalisateurs (notés MF) ont suscité l'intérêt des linguistes aussi bien du point de vue syntaxique que sémantique avec une certaine emphase sur leur valeur discursive. Par manque d'espace, nous ne présenterons pas ici les principales approches théoriques qui abordent le thème. Citons toutefois les travaux de Zubizarreta (1998) et de Birch *et al.* (2000) pour une introduction générale sur la focalisation, ceux de Hoeksema et Zwarts (1991) sur les adverbes focalisateurs, en particulier, et celui de König (1991) dans une perspective comparative. Costa (2008) fait référence aux différentes études concernant les adverbes focalisateurs portugais. Azpiazu Torres (1998) et Kovacci (1999), entre autres, proposent une introduction au thème pour l'espagnol.

Pour cette recherche, nous prenons appui sur le cadre théorique et méthodologique du lexique grammairal, notamment sur Maurice Gross (1986) et sur Molinier et Levrier (2000) dont nous avons suivi le classement syntactico-sémantique. Les adverbes focalisateurs (MF) sont des cas spéciaux d'adverbes modifieurs intégrés à la proposition et, généralement, ils sont rattachés à l'un des constituants majeurs de la phrase :

[1] Jean lit *principalement* des romans policiers.

Nous identifions sous la dénomination MF la construction de *principalement* illustrée dans [1] pour sa portée sur le constituant, ici un groupe (ou syntagme) nominal (SN), *des romans policiers* ; la phrase [1] implique qu'il existe un ensemble d'objets définis par *Jean lit* mais qu'un seul sous-ensemble a été retenu.

Cet article est structuré comme suit : nous définirons, dans un premier temps (voir 1.1), la notion de focalisation et les adverbes focalisateurs en fonction de leurs principales propriétés linguistiques ; nous montrerons que les différentes constructions/structures d'un même adverbe (voir 1.2) peuvent aboutir à des représentations syntaxiques et sémantiques de la phrase très dissemblables (voir 1.3). Nous soutiendrons ensuite l'importance de l'approche contrastive comme méthode pour la découverte des propriétés linguistiques (voir 1.4). Par la suite (voir 2), nous présenterons deux expériences pour la tâche de désambiguïsation et classification des adverbes MF. Puis (voir 3), nous discuterons les résultats de ces expériences (voir 3.1) afin d'évaluer l'apport des résultats obtenus dans la détermination des dépendances syntaxiques entre ces adverbes et les constituants sur lesquels ils portent (voir 3.2). L'article se termine (voir 4) par un résumé de nos résultats et les perspectives qui s'ouvrent à la suite de cette recherche.

1 Prolégomènes

1.1 Définition

Les adverbes focalisateurs sont des modifieurs internes à la proposition qui portent sur un groupe nominal (SN) ou un complément prépositionnel (SP). Les adverbes focalisateurs peuvent aussi bien porter sur une phrase subordonnée : Jean voulait principalement faire cela. Ils sont définis formellement par les propriétés syntaxiques suivantes (Molinier et Levrier 2000) : (i) ils ne peuvent pas être détachés en tête de phrase négative [1a] ; (ii) ils ne peuvent pas être extraits avec *c'est... que* [1b] ; (iii) mais ils peuvent être extraits avec le constituant principal sur lequel ils portent [1c] :

[1a] *Principalement, Jean ne lit pas des romans policiers.

[1b] *C'est principalement que Jean lit des romans policiers.

[1c] C'est principalement des romans policiers que Jean lit.

Ces adverbes portent, donc, sur un constituant majeur de la proposition (ici le complément d'objet direct) et non pas sur le verbe de la phrase.

Les adverbes focalisateurs constituent une sous-classe d'éléments lexicaux relativement petite, qui comprend non seulement les adverbes dérivés en *-ment*, comme *principalement*, mais aussi d'autres adverbes, comme *surtout*, y compris quelques adverbes composés, parfois en relation morphosyntaxique avec l'adverbe en *-ment*, comme *en particulier*.

1.2 Ambigüité

Comme c'est le cas pour bon nombre d'unités lexicales, la plupart des adverbes MF apparaissent dans plusieurs constructions/structures, où ils présentent alors un sens différent et portent sur différents constituants de la proposition. Observons, par exemple, l'adverbe *justement* dans les phrases suivantes (exemples tirés du corpus, voir 2.2) :

[2] Justement avant le débarquement, il y a eu concentration des forces françaises. (MF)

[3] Je vous cherchais justement. (PC)

[4] L'État justement condamné pour faute lourde. (MS)

Dans la phrase [2], l'adverbe joue un rôle proprement de focalisation portant sur le complément circonstanciel de temps. Toutefois, dans les autres phrases, l'adverbe fonctionne différemment; ainsi, en [3], *justement* est un adverbe de phrase conjonctif (PC), qui établit avec le discours précédent une relation sémantique précise (explication-conséquence); ce statut conjonctif est évident du fait qu'il ne peut pas apparaître, dans ce sens, en début absolu de discours. En [4], il s'agit d'un adverbe de manière orienté vers le sujet (MS), car l'adverbe peut être mis en relation avec une phrase adjectivale à sujet humain; comme il s'agit d'une phrase passive, cette relation s'établit avec l'agent (qui a été ici réduit à zéro): *L'État a été condamné [par le Tribunal] # Le Tribunal a été juste (dans la condamnation de l'État)*.

Nous avons adopté la classification des adverbes qui a été proposée par Molinier & Levrier (2000) pour le français. Notons toutefois que nous divergeons pour ce qui est du critère de définition des adverbes quantifieurs MQ. En fait, nous pensons qu'ils peuvent aussi être considérés comme quantifieurs lorsqu'ils portent sur un déterminant quantifieur (p. ex. un numéral) d'un SN, comme nous pouvons le voir dans l'exemple [5]; pour ces auteurs, il s'agirait d'une construction de type MF :

[5] Jean mesure exactement 1m80 (MQ)

1.3 Représentation sémantique

En traitement automatique du langage (TAL), lors d'une analyse syntaxique de surface (appelée aussi *shallow parsing* ou *chunking*), souvent la distinction entre

plusieurs types (ou classes) d'adverbes n'a pas de conséquences importantes. L'adverbe est normalement étiqueté comme un syntagme adverbial (noté *SADV*). Par exemple, (1) serait noté :

[1d] [Jean]_{SN} [lit]_V [principalement]_{SADV} [des romans policiers]_{SN}

L'analyse syntaxique profonde (*deep parsing*), c'est-à-dire, l'établissement des relations syntaxiques entre les constituants de la phrase (sujet, complément d'objet, modifieur, déterminant, etc.), est fréquemment une tâche subséquente à cette analyse de surface préliminaire. En vue de la relation particulière que les adverbes focalisateurs établissent avec le constituant sur lequel ils portent, il est question d'associer d'une manière adéquate l'adverbe au groupe nominal/complément prépositionnel et non pas au verbe de la phrase. Ceci peut être fait, comme, par exemple, dans le cadre du *xip*, Xerox Incremental Parser, (Ait-Mokhtar *et al.* 2002), tel qu'il a été adapté pour le portugais par Mamede *et al.* (2007), par l'établissement de dépendances entre les constituants identifiées par l'analyseur (*chunker*) ; dans notre exemple, on aurait, donc, de façon simplifiée, l'ensemble de dépendances suivantes :

TOP (lit)
 SUBJ (lit, Jean)
 OBJ (lit, romans policiers)
 DET (romans policiers, des)
 MODIF (romans policiers, principalement)

Fig. 1 : Dépendances syntaxiques extraites de la phrase : *Jean lit principalement des romans policiers*

Néanmoins, ceci n'est pas seulement fonction des items lexicaux qui sont classés comme des adverbes focalisateurs : la position relative de l'adverbe par rapport à d'autres constituants de la phrase, l'ambiguïté avec d'autres constructions/structures du même adverbe, et d'autres facteurs encore jouent un rôle qui rend non triviale l'identification et l'attribution de la dépendance syntaxique adéquate à ces mots. Ainsi, si on reprend les exemples [2] à [4] ci-dessus, on aurait les dépendances suivantes ; par simplicité, on indique seulement la portée de l'adverbe sur le mot tête (HEAD) du constituant (*chunk*) identifié dans l'exemple et les dépendances qui lui sont attachées (Fig. 2) :

[2] [Justement]_{ADVP} [avant le débarquement]_{PP}, il y a eu concentration des forces françaises (MF)

PP (débarquement, avant)

HEAD (débarquement)

DETD (débarquement, le)

ADVP (justement)

HEAD (justement)

MODIF (débarquement, justement)

[3] Je vous [cherchais]_{VF} [justement]_{ADVP} (PC)

VF (cherchais)
 HEAD (cherchais)
 MODIF (cherchais, justement)

[4] L'État [justement]_{PADV} [condamné]_{VCPART} pour faute lourde (MS)
 VCPART (condamné)
 HEAD (condamné)
 MODIF (condamné, justement)

Fig 2 : Dépendances syntaxiques et ambiguïté

1.4 Approche contrastive

Dans bon nombre de cas, il existe des adverbes focalisateurs équivalents dans différentes langues, tout particulièrement dans des langues aussi proches que le portugais et l'espagnol :

O João lê principalmente romances policiais.

Juan lee principalmente novelas policíacas.

Mais, parfois, nous pouvons trouver des faux amis ; ainsi, l'adverbe *en particulier/particulièrement* fonctionne comme un MF en espagnol (ES) :

Juan habló de esto en particular/particularmente. (ES)

Jean a parlé de cela en particulier/particulièrement.

Alors qu'en portugais, les formes correspondantes peuvent aussi présenter une construction MV, équivalente à *en privé/privément* :

O João falou em particular/particularmente com a Maria. (PT)

Jean a parlé avec Maria en privé/privément

D'autre part, nous avons recensé dans notre corpus une construction MV pour l'adverbe espagnol *inclusivamente*, « inclusivement » :

Seguimos pensando que hay que trabajar inclusivamente/MV, que todos tenemos derecho al Arte y a engrandecer nuestra cultura.

Cet emploi (apparemment récent) semble un emprunt de l'anglais comme nous pouvons le voir dans l'exemple ci-dessous :

As a matter of priority that school staff always work inclusively with the parents/carers of their students « en priorité ce personnel de l'école travaille inclusivement avec les parents/tuteurs de leurs étudiants ».

Finalement, nous avons aussi relevé des adverbes focalisateurs spécifiques à chaque langue, notamment *designadamente* et *nomeadamente* « particulièrement », qui n'existent qu'en portugais, et *sencillamente*, « simplement » qui est exclusif de l'espagnol. En outre, il faut signaler que les focalisateurs *mormente* (PT) et *mayormente*

(ESP) « spécialement », semblent ne plus être à la mode, malgré leur emploi dans les deux langues.

Malgré ces différences, ils se comportent de façon remarquablement semblable à celle rapportée pour le français (Molinier et Levrier 2000), ce qui appelle une étude comparative. Nous pensons donc qu'une approche contrastive peut souligner les différents sens et contextes d'emploi et, par là même, permettre de cerner les phénomènes linguistiques impliqués.

2 Méthodes

Le but de cet article est d'évaluer l'importance de l'apport de l'information syntactico-sémantique vis-à-vis des adverbes focalisateurs pour la tâche de délimitation de groupes nominaux en portugais et en espagnol. Pour ce faire, il faut préalablement désambigüiser ces adverbes. Nous avons adopté une approche de désambigüisation basée sur des règles afin de distinguer les différentes constructions/structures de ces adverbes et les sens correspondants. Finalement, nous avons procédé à l'évaluation de l'analyse syntaxique des groupes nominaux corrigée par l'apport de cette information syntactico-sémantique.

2.1 Catalogue de sens et des structures syntaxiques

Nous avons recensé un catalogue de sens et de structures syntaxiques des adverbes focalisateurs pour chacune des langues comme nous pouvons le voir dans la Table 1 et la Table 2. Les différents sens correspondent d'une manière générale aux différentes classes syntactico-sémantiques établies en suivant les critères proposés par Molinier et Levrier (2000) avec quelques adaptations moindres. Nous indiquons aussi le nombre d'adverbes repérés dans le corpus d'entraînement (voir 2.2 ci-dessous).

| Français | Portugais | Classes | Count |
|-------------------|-------------------------|------------|-------|
| approximativement | <i>aproximadamente</i> | MF | 237 |
| basiquement | <i>basicamente</i> | MF, MQ | 323 |
| concrètement | <i>concretamente</i> | MF | 222 |
| notamment | <i>designadamente</i> | MF | 10 |
| spécialement | <i>especialmente</i> | MF, MQ | 437 |
| spécifiquement | <i>especificamente</i> | MF, MQ, MV | 377 |
| essentiellement | <i>essencialmente</i> | MF, MQ | 109 |
| exactement | <i>exactamente</i> | MF, MQ, MS | 459 |
| exclusivement | <i>exclusivamente</i> | MF | 443 |
| fondamentalement | <i>fundamentalmente</i> | MF, MQ | 82 |
| inclusivement | <i>inclusivamente</i> | MF | 340 |
| justement | <i>justamente</i> | MF, PC, MS | 388 |
| seulement | <i>meramente</i> | MF | 224 |
| surtout | <i>[mormente]</i> | MF | 106 |
| notamment | <i>nomeadamente</i> | MF | 247 |
| particulièrement | <i>particularmente</i> | MF, MQ, MV | 190 |
| personnellement | <i>pessoalmente</i> | MF, PS | 375 |

| | | | |
|----------------|------------------------|----------------|-------|
| pratiquement | <i>praticamente</i> | MF, PS, MQ, MV | 710 |
| précisément | <i>precisamente</i> | MF, PC, MS, MV | 446 |
| principalement | <i>principalmente</i> | MF | 331 |
| respectivement | <i>respectivamente</i> | MF | 195 |
| simplement | <i>simplesmente</i> | MF, PC, PS, MS | 419 |
| seulement | <i>somente</i> | MF | 548 |
| uniquement | <i>unicamente</i> | MF | 390 |
| | | Total | 7 608 |

Table 1 : Adverbes MF en portugais

| Français | Espagnol | Classes | Count |
|-------------------|-------------------------|--------------------|--------|
| approximativement | <i>aproximadamente</i> | MF | 3 270 |
| basiquement | <i>básicamente</i> | MF, MQ | 380 |
| concrètement | <i>concretamente</i> | MF | 721 |
| notamment | – | – | – |
| spécialement | <i>especialmente</i> | MF, MQ | 429 |
| spécifiquement | <i>especificamente</i> | MF, MQ, MV | 233 |
| essentiellement | <i>esencialmente</i> | MF, MQ | 470 |
| exactement | <i>exactamente</i> | MF, MQ, MS | 401 |
| exclusivement | <i>exclusivamente</i> | MF | 387 |
| fondamentalement | <i>fundamentalmente</i> | MF, MQ | 22 |
| inclusivement | <i>inclusivamente</i> | MF, MV | 131 |
| justement | <i>justamente</i> | MF, PC, MS | 380 |
| seulement | <i>meramente</i> | MF | 179 |
| surtout | <i>mayormente</i> | MF | 284 |
| notamment | – | – | – |
| particulièrement | <i>particularmente</i> | MF, MQ | 434 |
| personnellement | <i>personalmente</i> | MF, PS | 1 468 |
| pratiquement | <i>prácticamente</i> | MF, PS, MQ, MV | 837 |
| précisément | <i>precisamente</i> | MF, PC, MS | 430 |
| principalement | <i>principalmente</i> | MF | 387 |
| respectivement | <i>respectivamente</i> | MF | 949 |
| simplement | <i>sencillamente</i> | MF | 355 |
| simplement | <i>simplesmente</i> | MF, PC, PS, MS, MQ | 361 |
| | | Total | 13 169 |

Table 2 : Adverbes MF en espagnol

2.2 Corpus d'entraînement

Nous avons construit un petit corpus à partir de phrases extraites arbitrairement de l'internet en utilisant le système de recherche *WebCorp* (www.webcorp.org.uk). Nous avons aussi pris en compte les options suivantes: les concordances ont été

obtenues à partir du moteur de recherche *Google*, de façon insensible aux majuscules et aux minuscules, dans tous les domaines textuels, en récupérant une phrase complète de texte par concordance, une concordance par site internet, en format KWIC ; les sites internet, les liens et les adresses électroniques ont été exclus. Pour chaque domaine (.pt et .es) nous avons téléchargé jusqu'à 500 occurrences de chaque adverbe. Le corpus portugais a 210 K mots simples (dont 26,8 K différents) tandis que le corpus espagnol présente 361,3 K mots (dont 37,9 K différents).

Une sélection de 10 occurrences de chaque adverbe a été extraite de façon aléatoire du corpus d'entraînement avant la recherche des motifs syntactiques pour la construction des règles de désambiguïsation (voir 2.3 ci-dessous). Ce corpus constituera le corpus d'expérimentation (voir 2.4 ci-dessous). Nous avons retiré un exemple toutes les dix occurrences ; nous n'avons, donc, retenu que les exemples sans fautes d'orthographe, qui n'étaient pas des titres, et où les adverbes étaient inclus dans une phrase complète, ce qui n'était pas toujours le cas dans le corpus extrait directement d'internet.

2.3 Règles de désambiguïsation

Puis, nous avons déterminé des règles linguistiques permettant de désambigüiser le sens des mots. Les règles ont été implémentées et appliquées à des corpus en utilisant des transducteurs à états finis (FST) avec la plateforme de développement linguistique UNITEX (Paumier, 2003, 2008). La Fig. 3 illustre un de ces graphes de désambiguïsation, notamment celui de l'adverbe *simplemente* (« simplement ») de l'espagnol :

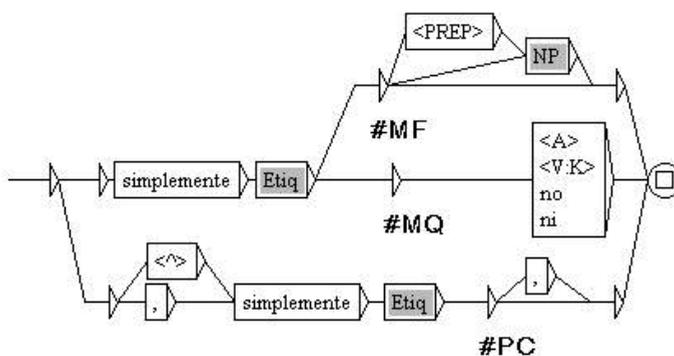


Fig. 3 : FST pour la désambiguïsation de l'adverbe espagnol *simplemente*

Ce FTS se lit comme suit : par défaut, l'adverbe *simplemente* est classé comme MF devant un groupe nominal ou prépositionnel (en faisant appel à des sous-graphes) ; devant un adjectif <A> ou un participe passé <V: K> ou, encore, les adverbes de négation *no* et *ni*, il est classé comme MQ ; finalement, en début de phrase <^> ou entre virgules, *simplemente* est classé comme adverbe de phrase, PC.

Nous remarquons que ces règles ont été d'abord élaborées pour l'espagnol et ensuite appliquées au corpus portugais, sans d'autres changements que l'adaptation à l'orthographe et l'inclusion/exclusion des cas différents dans chaque langue. Nous avons voulu ainsi estimer la portabilité de la grammaire de désambiguïsation d'une langue à l'autre. En fait, ceci a eu un impact sensible dans les résultats de la tâche de classification, comme on le verra dans la section 3.

Dans un deuxième temps, nous avons construit un second ensemble de FST, mais en remplaçant les adverbes par les étiquettes correspondantes : <ADV + MF>, <ADV + MQ>, etc. (Fig. 4). L'objectif de ce deuxième ensemble de transducteurs est de vérifier s'il y a dégradation des résultats des grammaires quand elles sont généralisées au niveau de la classe d'adverbes.

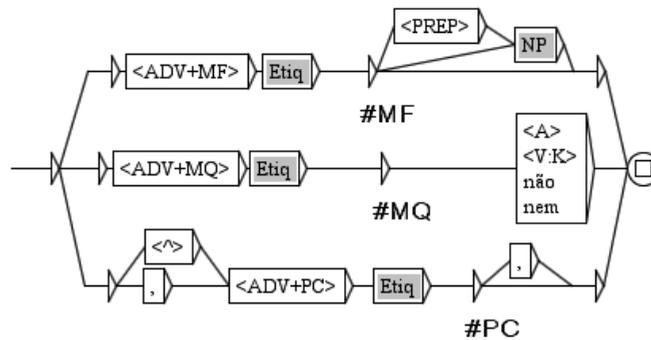


Fig. 4 : FST généralisé à partir de la grammaire de désambiguïsation construite pour l'adverbe simple

Nous avons aussi utilisé une grammaire simple pour la délimitation de groupes nominaux. L'alphabet de cette grammaire prend seulement les catégories basiques (nom, pronom, adjectif, adverbe et préposition), mais pour les déterminants, nous avons aussi eu recours au symbole <NB>, traditionnellement utilisé pour les nombres. Le groupe nominal admet aussi comme tête des suites de noms en majuscules, pour les noms propres. La grammaire utilise les dictionnaires du portugais et de l'espagnol distribués avec le logiciel Unitex.

2.4. Corpus d'expérimentation

Comme il a déjà été dit au point 2.2 *in fine*, nous avons sélectionné 10 concordances de chaque adverbe, pour chaque langue. Cette sélection a été étiquetée manuellement en fonction des différents sens/constructions présents dans chaque exemple, et cette annotation a été faite séparément par chacun des deux auteurs. Puis les corpus annotés ont été mis en commun. Dans un premier temps, l'accord inter-annotation était déjà très élevé (85 %) mais pratiquement tous les cas divergents ont été résolus après discussion. Ces deux corpus annotés constituent notre *golden standard*. La Table 3 présente la distribution des divers types d'adverbes d'après la classification.

| Corpus | MF_T1 (%) | MF_T2 (%) | MQ | MS | MV | PS | PC | Mx | Total |
|-----------|-----------|------------|----|----|----|----|----|----|-------|
| Espagnol | 88 (0,68) | 188 (0,82) | 24 | 3 | 2 | 2 | 8 | 3 | 230 |
| Portugais | 80 (0,67) | 200 (0,83) | 23 | 5 | 7 | 2 | 1 | 2 | 240 |

Table 3 : Distribution des types d'adverbes dans les corpus

Les corpus ont été sous divisés en deux parties chacun :

- Text1 contient seulement les exemples avec des adverbes potentiellement ambigus (ES : 130 phrases, PT : 120 phrases) ;
- Text2 comprend toutes les phrases des corpus d'expérimentation (ES : 230 phrases, PT : 240 phrases).

Le but de cette division était d'évaluer l'amélioration que les règles de désambiguïsation peuvent introduire dans la tâche de classification. Dans la Table 3 ci-dessus, cette distinction est réfléchiée dans la distribution des adverbes de focalisation : MF_T1 ou MF_T2, selon qu'il s'agit du corpus Text1 ou Text2, respectivement.

Indépendamment du sous-corpus, on remarque, dans la Table 3, que la plupart des adverbes peuvent se classer comme MF ; dans Text1 les MF correspondent à 67 ou 68 %, tandis que dans Text2 ce pourcentage augmente jusqu'à 82 ou 83 %, en espagnol et en portugais, respectivement. La deuxième classe la plus significative est MQ, avec environ 18 ou 19 % dans Text1 et 10 % dans Text2.

Dans le corpus espagnol, trois phrases n'ont pas pu être classifiées de façon précise du fait que les deux annotateurs ne se sont pas mis d'accord. Seulement deux phrases de ce type ont été repérées dans le corpus portugais. Nous avons employé dans ces cas une étiquette spécifique (Mx).

2.5 Bordures statistiques supérieure (*ceiling*) et inférieure (*baseline*)

En comptant les cas que nous n'avons pas pu classer (Mx), nous avons estimé une bordure statistique supérieure (*ceiling*) de 0,987 pour la tâche de classification dans le corpus espagnol et 0,992 pour le portugais ; ce qui implique, d'après ces corpus, qu'il y a un maximum de cas pour lesquels il est possible de proposer automatiquement une classification, car là où les annotateurs humains n'arrivent pas à se mettre d'accord, aucun système ne devrait être capable non plus d'arriver à une classification fiable. D'un autre côté, comme il est question ici des adverbes dont l'emploi comme focalisateur est clairement prédominant, nous avons calculé la bordure statistique inférieure (*baseline*) comme taux de précision qui aurait été obtenu si toutes les occurrences de ces adverbes avaient été classées comme MF pour chaque sous-corpus. Cette *baseline* est différente selon qu'il s'agit du corpus Text1 ou Text2 (Table 4) :

| Corpus | Espagnol | | Portugais | |
|----------|----------|-------|-----------|--------|
| | Text1 | Text2 | Text1 | Text 2 |
| Ceiling | 0,987 | | 0,992 | |
| Baseline | 0,68 | 0,82 | 0,67 | 0,83 |

Table 4 : Ceiling et Baseline

3 Résultats et discussion

Comme nous l'avons dit plus haut, le but de cet article est d'évaluer l'importance de l'apport de l'information syntactico-sémantique par rapport aux adverbes focalisateurs pour la tâche de délimitation de groupes nominaux en portugais et en espagnol. Dans cette visée, nous avons procédé à deux ensembles d'expériences.

En premier lieu, nous avons évalué la tâche de classification des adverbes, en utilisant les grammaires de désambiguïsation. Cela s'est fait en deux étapes: dans l'expérience 1 (Exp1) nous avons appliqué les grammaires lexicalisées, c'est-à-dire, construites autour des adverbes, comme il a été exemplifié dans les Fig. 1 et Fig. 2. Postérieurement, dans l'expérience 2 (Exp2) nous avons appliqué des grammaires identiques, mais en remplaçant les adverbes par leurs étiquettes. L'objectif de cette deuxième étape est de déterminer la dégradation des résultats quand une grammaire lexicalisée est généralisée à la classe de l'adverbe cible.

3.1 Tâche de classification

Les résultats préliminaires sont présentés dans la Table 5.

| | Résultats espagnol | | | | | Résultats portugais | | | | |
|------------|--------------------|------|------|------|------|---------------------|------|------|------|------|
| | P | R | F | B | G | P | R | F | B | G |
| Text1_Exp1 | 0,80 | 0,78 | 0,79 | 0,68 | 0,13 | 0,69 | 0,67 | 0,68 | 0,67 | 0,03 |
| Text1_Exp2 | 0,83 | 0,80 | 0,82 | 0,68 | 0,15 | 0,70 | 0,70 | 0,70 | 0,67 | 0,04 |
| Text2_Exp1 | 0,89 | 0,79 | 0,83 | 0,82 | 0,07 | 0,82 | 0,68 | 0,75 | 0,83 | 0,01 |
| Text2_Exp2 | 0,91 | 0,85 | 0,87 | 0,82 | 0,09 | 0,84 | 0,81 | 0,83 | 0,83 | 0,01 |

P = précision, R = couverture (recall), F = mesure F ($2PR/P + R$), B = baseline, G = gain

Table 5 : Résultats préliminaires de la tâche de classification

Remarquons d'abord que les grammaires de désambiguïsation produisent des résultats bien meilleurs que la *baseline* pour le corpus de l'espagnol. Les résultats montrent que si l'on part d'une bordure statistique inférieure (*baseline*) relativement faible (0,68) les règles de désambiguïsation améliorent d'une manière significative la performance de la classification (+0,13 et +0,15, pour chaque expérience Exp1 et Exp2, respectivement), lorsque seuls les adverbes ambigus sont impliqués (Text1). Ceci est naturel car environ 43 % du corpus est constitué d'exemples MF

non ambigus. Lorsque tous les adverbes de focalisation sont considérés (Text₂), l'amélioration n'est que légèrement meilleure (+0,07 et +0,09, dans Exp₁ et Exp₂, respectivement) alors que la bordure statistique inférieure pour ce sous-corpus est également plus élevée (0,82). La précision globale, la couverture et la mesure F ($F = 2PR/P + R$) est similaire dans les deux corpus et dans les deux expériences, avec des valeurs plus élevées dans l'expérience 2 en utilisant le corpus Text₂.

Les résultats de la tâche de classification sur le corpus portugais sont plus faibles que ceux de l'espagnol. Ils sont même inférieurs à la *baseline* pour le Text₂, quand tous les adverbes, y compris les non ambigus sont pris en compte. Cela pourrait être dû au fait que les grammaires ont été développées d'abord pour l'espagnol et appliquées ensuite sans changements (sauf pour des questions d'orthographe) au corpus portugais. Nous avons, donc, étudié la matrice de confusion, pour mieux comprendre les raisons de cette performance. Pour des raisons d'espace, nous ne pouvons pas la montrer ici. Il découle de l'analyse de ces résultats que la plupart des erreurs de classification sont dues à la confusion entre MF et MQ et à un usage assez différent dans le corpus des virgules avec les adverbes PC et PS.

Il est intéressant de remarquer que si l'on remplace les étiquettes des mots (de l'expérience₁) par leurs étiquettes lexicales, selon l'étiquette de sortie désirée (expérience₂) les résultats ne changent presque pas, en fait ils s'améliorent légèrement mais systématiquement (de 0.01 à 0.02). Ceci peut s'interpréter par la distribution semblable des MF et suggère que les règles de désambiguïsation pourraient être tirées indépendamment des adverbes particuliers qui les ont inspirées.

3.2 Tâche d'identification des dépendances syntaxiques

Pour la tâche d'identification des dépendances syntaxiques, nous avons utilisé la grammaire locale des groupes nominaux (SN) pour les délimiter. La même grammaire nous a aussi servi pour identifier les compléments prépositionnels (SP), après de petits ajouts. Nous avons limité notre évaluation aux adverbes MF et MQ, à cause de leur importance quantitative dans les deux corpus et aussi parce que ce type de quantifieurs s'analysent souvent comme faisant partie du groupe nominal. Nous ne considérons ici que les cas où la tâche de classification a réussi. L'évaluation a été faite sur les corpus qui contiennent seulement des adverbes ambigus (Text₁) après l'application des grammaires de désambiguïsation lexicalisées (Exp₁).

Les résultats obtenus sont présentés dans la Table 6. Pour le corpus Text₁ de chaque langue, on présente le total de cas MF et MQ bien classés, suivi de la précision (P) obtenue dans la tâche d'identification des dépendances syntaxiques :

| Portugais | | | | Espagnol | | | |
|-----------|--------|-------|--------|----------|--------|-------|--------|
| Total | P (MF) | Total | P (MQ) | Total | P (MF) | Total | P (MQ) |
| 79 | 0,87 | 8 | 0,88 | 83 | 0,90 | 18 | 0,89 |

Table 6 : Tâche d'identification des dépendances syntaxiques dans les corpus avec des adverbes ambigus (Text1). Total = total de cas correctement classifiés e P = précision, pour chaque classe.

Pour conclure cette communication, nous ferons quelques commentaires sur le corpus et l'évaluation. Il convient de souligner que le nombre d'occurrences pour chaque construction du même adverbe dans le corpus d'expérimentation peut ne pas représenter adéquatement sa distribution normale. En particulier, nous avons tâché d'inclure au moins un exemple de chaque construction entre les exemples du même adverbe sélectionné pour ce corpus, ce qui affectait la performance et diminuait le taux de précision. Finalement, nous sommes conscients que, en raison du nombre réduit d'exemples dans le corpus, la limite supérieure est probablement trop optimiste. En fait, la plupart des distinctions tiennent à des différences sémantiques subtiles marquées par de nombreux facteurs difficiles à contrôler correctement dans un corpus de base. À plus grande échelle, cela est susceptible d'entraîner beaucoup plus de désaccords d'annotation ce qui ferait diminuer la limite supérieure jusqu'à un scénario plus réaliste.

Conclusions et perspectives

Il reste, bien sûr, des aspects qui peuvent être améliorés, notamment, la découverte de nouvelles règles de désambiguïsation et son ajout à la grammaire. D'un autre côté, il serait très intéressant de comparer ces résultats avec des méthodes d'apprentissage automatique. Pour le faire, un corpus annoté plus large devrait être construit pour l'apprentissage et, pour l'évaluation, on devrait être capable de disposer d'un corpus annoté avec une distribution des adverbes plus naturelle.

Bibliographie

- Ait-Mokhtar (S.), Chanod (J.-P.), Roux (Claude), 2002 : « Robustness beyond shallowness: incremental dependency parsing », dans *Natural Language Engineering* 8 (2/3), p. 121-144.
- Azpiazu Torres (S.), 1998 : « Acerca de la focalización adverbial », dans *Interlingüística* 9, p. 31-36.
- Baptista (J.), Català (D.), 2009 : *Disambiguation of focus Adverbs in Portuguese and Spanish*, BULAG 33, p. 31-37.
- Birch (S.), Albrecht (J.), Myers (J.), 2000 : « Syntactic Focusing Structures Influence Discourse Processing. », dans *Discourse Processes* 30-3, p. 285-304.
- Borgonovo (C.), 2003 : « Mood and Focus », dans *Studies in the Theory and History of Linguistic Science* series 4, 245, Amsterdam, p. 17-30.
- Burgess (J.), Etherington (S.), 2002 : « Focus on grammatical form: explicit or implicit? », dans *System*, 30/4, p. 433-458.
- Costa (J.), 2008 : *O Advérbio em Português Europeu*, Lisboa, Colibri.

- Gross (M.), 1986 : *Grammaire transformationnelle du français – III. Syntaxe de l'adverbe*, Paris, ASSTRIL.
- Hoeksema (J.), Zwarts (F.), 1991 : « Some Remarks on Focus Adverbs », dans *Journal of Semantics* 8, p. 51-70.
- König (E.), 1991 : *The Meaning of Focus Particles: A Comparative Perspective*, Oxford, Routledge.
- Kovacci (O.), 1999 : « El adverbio », dans Bosque (I.), Demonte (V.), dir., *Gramática Descriptiva de la Lengua Española*, Madrid, Espasa Calpe.
- Mamede (N.), Baptista (J.), Vaz (P.), Hagège (C.), 2007 : *Nomenclature of chunks and dependencies in Portuguese XIP grammar 2.1*. (Internal Report), Lisboa, L2F/INESC ID Lisboa.
- Molinier (C.), Levrier, (Fr.), 2000 : *Grammaire des adverbes, Description des formes en -ment*, Genève Paris, Droz.
- Paumier (S.), 2003 : *De la reconnaissance de formes linguistiques à l'analyse syntaxique*, PhD Thesis, Paris, Université de Marne-la-Vallée.
- Paumier (S.), 2008 : *Unitex 2.0 User's Manual*, Paris, Université de Marne-la-Vallée.
- Zubizarreta (M. L.), 1998 : *Prosody, Focus, and Word Order*, Cambridge, MTass., MIT Press.

Carla Serhan,
Département de langues vivantes et traduction,
Université de Balamand,
Tripoli,
Liban,
carla.serhan@balamand.edu.lb.

La traduction de la forme pronominale réfléchie en arabe et en anglais : révélation du rapport au corps

Résumé

Ne contenant pas de pronoms réfléchis, la langue arabe rend la forme pronominale réfléchie française par la dérivation verbale, le sémantisme verbal et le substantif *nafsun*. Les deux derniers procédés sont également utilisés par la langue anglaise : *nafsun* aura pour équivalent *self*. Ces procédés disparaissent dès qu'il s'agit du « réfléchi partiel » relatif à une partie du corps. Ainsi : « Il se lave les mains » ET « Il lave sa voiture », sont traduits de la même façon en arabe et en anglais. D'où l'étude de la question culturelle du rapport au corps, telle qu'elle est perçue dans les trois langues.

Mots-clés : verbe pronominal réfléchi ; traduction ; français ; arabe ; anglais ; rapport au corps.

Introduction

Le français au Proche-Orient et au Liban en particulier est en interaction avec deux autres langues principales : l'arabe et l'anglais. Le recours à ces trois langues est dans certains cas quotidien. Mais comme il est assez rare qu'une personne maîtrise les trois langues à la fois, ce passage est assuré par la médiatrice des langues et des cultures, à savoir la traduction.

Dans ce carrefour linguistique et culturel, une question retient notre attention : comment traduit-on la forme pronominale réfléchie dans les langues arabe et anglaise ? La question est légitime, car les verbes pronominaux n'existent en tant que tels, ni dans l'une ni dans l'autre de ces deux langues. Mais loin d'être intraduisible, cette forme pronominale se trouve rendue par plusieurs moyens qui, parfois, sont communs aux deux langues en question, et d'autres fois parfaitement distincts. Le fait que l'opération traduisante adopte un autre « tour » que celui du texte original, entraîne-t-il un changement significatif dans l'apport sémantique véhiculé par la forme pronominale française ?

Un élément révélateur nous aidera à répondre à cette question et par suite à avancer dans le cadre de cette étude, il s'agit du « réfléchi partiel » qui fait référence à une partie du corps humain.

Après une illustration à l'aide d'exemples révélateurs traduits, cette étude analysera, à partir de ces traductions, la question culturelle du rapport au corps, telle qu'elle est perçue dans les trois langues qui représentent trois visions différentes du monde. Cette analyse nous aidera à expliquer quelques convergences et divergences de nos trois langues. Ce qui nous éclairera sur certains points précis des modes de pensée adoptés par ces trois cultures.

En effet, le fait que le langage soit inséparable de la pensée n'a pas besoin d'être prouvé. Selon Noam Chomsky (1981 : 12), « les concepts exprimés et les distinctions établies dans l'usage courant du langage nous éclairent sur les schémas de pensée et l'univers du « sens commun » construits par l'esprit humain ». Le langage – y compris la langue définie par Saussure (1998 : 25) comme « sociale dans son essence » – est tout simplement, d'après Chomsky, (1981 : 12) « le miroir de l'esprit ».

Dans cette optique, examinons le cas particulier de l'expression de la notion de réfléchi à travers la traduction des verbes pronominaux.

1 Les procédés utilisés dans la traduction de la forme pronominale réfléchie

1.1 Bref aperçu de la forme pronominale

Nous allons dans ce qui suit présenter brièvement la forme pronominale. En effet, cette dernière se manifeste dans la langue française de deux façons différentes. La première correspond aux verbes dits essentiellement pronominaux qui se distinguent par le fait que « l'unité morphologique de voix pronominale fait partie du lexème verbal concerné [et] ne saurait commuter » (Touratier 1996 : 195). Nous emprunterons l'exemple de Riegel, Pellat, et Rioul (1999 : 255) pour illustrer nos propos « *s'enfuir*/**l'enfuir*/**enfuir quelqu'un*/**enfuis-le* ». La deuxième catégorie regroupe les verbes pronominaux « dont l'unité morphologique ne fait pas partie du lexème verbal » (Touratier 1996 : 196). On distingue alors la forme pronominale réciproque « Ils se sont tellement aimés ! », celle à sens passif « Les feuilles mortes se ramassent à la pelle », et la forme réfléchie « Il se regarde dans la glace ». C'est cette dernière réalisation qui intéresse notre étude.

« Syntactiquement et sémantiquement, la construction réfléchie d'un verbe se distingue de sa construction non réfléchie par le fait que le pronom complément y est coréférentiel au sujet. Dans *Paul s'admire*, le sujet est à la fois la source et le but du procès *admirer*, contrairement à *Paul l'admire* où admirateur et admiré sont deux personnes distinctes. » (Riegel, Pellat, et Rioul 1999 : 256)

Cette construction pronominale réfléchie n'existe pas en tant que telle dans les deux langues arabe et anglaise. D'autres moyens sont dans ce cas utilisés pour

exprimer la notion de « réfléchi » et traduire la forme pronominale. Nous allons tout de suite examiner ces différentes possibilités.

1.2 La dérivation verbale

Les langues sémitiques se caractérisent par une morphologie dérivationnelle exemplairement édifiée. Grâce au *qiyās* – qui est un ensemble de lois de dérivations qui ramène chaque mot à une forme ou schème connu dérivé lui-même de la forme nue, en l'occurrence la racine consonantique, le plus souvent trilitère – la langue arabe construit des formes verbales dérivées appelées aussi « verbes augmentés » (Kouloughli 1994 : 194). Ces formes sont au nombre de dix. Celles à sens « réfléchi » sont les cinquième, septième, huitième et dixième formes. Dans ce qui suit, nous allons illustrer chacun de ces quatre cas par deux exemples révélateurs, le premier représentera le verbe à la forme « simple », le second représentera la forme dérivée de ce même verbe. Chacun de ces exemples sera la traduction d'un énoncé français qui le précèdera. Il sera de même suivi d'une traduction anglaise. Tout en signalant que tous les exemples contenant un verbe pronominal réfléchi seront empruntés au récit de Maupassant « Une vie » et à ses deux traductions arabe et anglaise, dont les références figureront dans la liste bibliographique. La première traduction revient à Elie Khalil, et la seconde à Roger Pearson. Ces traductions seront suivies du numéro de page de la référence en question. Reste à signaler que la traduction arabe sera découpée en morphèmes et expliquée dans la ligne qui suit.

Voici la liste des abréviations utilisées dans ce but : Accompli : *Acc* ; Accusatif : *acc* ; Duel : *due* ; Féminin : *fém* ; Génitif : *gén* ; Inaccompli : *inac* ; Indicatif : *ind* ; Masculin : *masc* ; Nominatif : *nom* ; Personne : *pers* ; Pluriel : *plu* ; Singulier : *sing*, Subjonctif : *subj*.

1.2.1 La cinquième forme

La cinquième forme verbale *tafa'ala* « c'est la réfléchie-passive, [appelée aussi « la réfléchie intérieure »] de la deuxième forme *fa'ala* avec le préfixe *-ta* » (Blachère et Gaudefroy-Demombynes 1975 : 58) : *ta-zawwaʒa* = se marier, alors que : marier = *zawwaʒa*. Les exemples suivants illustreront nos propos :

[1] Il a marié sa fille à son neveu.

| | | | | |
|------|-----------------|------------------|----------------|-------------------|
| [1'] | <i>zawwaʒ-a</i> | <i>bnat-a-hu</i> | <i>li-bn-i</i> | <i>'aḥ-T-hi</i> . |
| | « marier-Acc | filles- acc-sa | à-fils-gén | frère-gén-son » |

[1"] *He married his daughter to his nephew.*

[2] Elles en ont toutes, des enfants, avant de se marier. (167)

| | | | | |
|------|------------------------------|-----------------------------|------------------|--|
| [2'] | <i>Kull-u-hu-nna</i> | <i>ju-nʒib-na</i> | <i>'aḥfāl-an</i> | <i>qabla 'an ja-tazawwaʒ-na (189)</i> |
| | « tout-nom- pers3-fém plu | pers3-enfan- ter-fém plu | enfants-acc | avant que pers3-se marier- fém plu. » |

[2"] *These girls all have children before they get married.* (126)

1.2.2 La septième forme

La septième forme verbale *'infa'ala* est « une réfléchie formée par la préfixation de *n*, précédée d'une voyelle légère d'attaque » (Blachère et Gaudefroy-Demombynes 1975 : 63) : *'in-sahaba* = « se retirer », alors que : « retirer » = *sahaba*. L'illustration de ce cas se fera à l'aide des exemples qui suivent :

[3] Il retira une somme d'argent.

[3'] *sahab-a mablag-an mina l-māl-i.*

« retirer-Acc somme-acc de le-argent-gén. »

[3"] *He withdrew some of money.*

[4] Mais la vieille fille [...] se retira dans une de ces maisons religieuses. (59)

[4'] *lākinna l-'ānis-a [...] nsahab-a-t 'ilā wāhidin min tilka l-bujūt-i d-dīnija-t-i.* (83)

« mais le-vieille fille-acc [...] se retirer-Acc-fém à un de ces le-maisons-gén le-religieux-fém-gén. »

[4"] *But the old maid [...] withdrew into one of those religious houses.* (42)

1.2.3 La huitième forme

La huitième forme verbale : *'ifta'ala* ou *fta'ala* « c'est la réfléchie-passive de la forme nue avec préfixation de *-ta* : mais le *-ta* se trouve infixé entre la première et la seconde consonne radicale du verbe. » (Blachère et Gaudefroy-Demombynes 1975 : 64) : comme [*i*] *rtafa'a* = « s'élever », alors que : « élever » = *rafa'a* :

[5] Il a élevé la voix.

[5'] *rafa'-a sawt-a-hu.*

« élever-Acc voix-acc-sa. »

[5"] *He raised his voice.*

[6] La voûte des nuées semblait s'élever, blanchir (12)

[6'] *'aqd-u l-ğujūm-i l-kaṭṭifa-t-i bada'-a
ja-rtafi'-u wa-ta-bjaḍḍ-u qubbat-u l-falak-i* (36)

« voûte-nom le-nuages-gén le-épais-fém-gén commence-Acc pers3-s'élever-ind et-pers3 fém-blanchir-ind voûte-nom le-ciel-gén »

[6"] *The clouds having now parted, the azure reaches of the sky appeared.* (8)

1.2.4 La dixième forme

La dixième forme verbale : *'istaf'ala* « c'est la réfléchie passive de la quatrième forme [*af'ala*] avec addition de *t*. Mais ici le préfixe de la quatrième forme, au lieu d'être *hamza*, a conservé l'ancien préfixe *-s*. » (Blachère et Gaudefroy-Demombynes 1975 : 69) :

comme *'istadāra* = « se tourner », alors que : « tourner » = *'adāra*. Deux exemples illustreront cette différence :

[7] Il tourna le ventilateur dans sa direction.

[7'] *'adār-a* *l-mirwaḥat-a* *bi-ttiḏāh-i-ha*.
« tourner-Acc le-ventilateur-acc dans-direction-gén-pers3 fém. »

[7"] *He pointed the fan towards her.*

[8] Il se tourna vers elle avec un visage mécontent. (99)

[8'] *'istadār-a naḥwa-h-a bi-waḏh-in ḡajr-i rāḍ-in*. (120)
« se tourner-Acc vers-pers3-fém avec-visage-gén non-gén satisfait-gén. »

[8"] *He turned towards her with a look of annoyance on his face.* (73)

1.3 Le sémantisme verbal

Un deuxième procédé est utilisé dans le même but, il s'agit des verbes qui véhiculent en eux-mêmes le trait sémantique « réfléchi ». L'exemple suivant est assez révélateur :

[9] puis il se tut par crainte d'en trop lâcher. (170)

[9'] *tumma samat-a ḥawf-an min 'an ju-flit-a l-kalām-u min-h-u* (191)
« puis se taire-Acc peur-acc de que pers3-échapper-subj le-parole-nom de-pers3-masc. »

[9"] *And then he paused for fear of giving too much away.* (127)

Signalons que le sémantisme verbal s'avère un moyen de grande fréquence dans la traduction anglaise de la forme pronominale réfléchie, preuve en est l'utilisation des verbes : *withdrew*, *pointed*, et *parted* dans les exemples précédemment cités.

1.4 *Nafsun* et... *self*

Un troisième procédé est utilisé pour rendre la forme pronominale réfléchie, il s'agit du substantif *nafsun*. Considérons l'exemple suivant :

[10] Elle se trouvait, en une seconde, délivrée, apaisée, heureuse. (163)

[10'] *waḏad-a-t nafsahā bi-laḥḏat-in ṭaliqa-t-an hādī'a-t-an saṭda-t-an*. (185)
« trouver-Acc-fém elle-même en-seconde-gén libre-fém-acc calme-fém-acc heureux-fém-acc. »

[10"] *In a single instant she found herself delivered, assuaged, happy.* (123)

D'après la grammaire de Blachère, en arabe « il n'y a pas de pronom réfléchi. L'équivalent du pronom français « moi-même », « toi-même », etc., est le substantif *âme*, *personne* : *nafsun*, construit avec le pronom affixe : *nafsī*, *nafsuka*, *nafsuki*, etc. » (1975 : 203). La langue anglaise use de *himself*, *herself*, etc. pour traduire la forme pronominale dans ce genre de cas.

La classification des différentes traductions de la forme pronominale s'est peut-être plus alignée sur les données de la langue arabe que sur celles de la langue anglaise. En effet, il n'est pas toujours aisé de procéder à la comparaison de trois langues, il vaut mieux choisir une voie et la suivre pour éviter le décousu et/ou la répétition. Cette voie a été pour nous celle de la langue arabe dont la morphologie jouit d'« une unité et [d']une sûreté de classement très pédagogique » (Blachère et Gaudefroy-Demombynes 1975 : 13-14). Les premières conclusions montrent que la morphologie joue un rôle très important dans la traduction arabe de la forme pronominale réfléchi, alors qu'elle n'a pratiquement aucun rôle à jouer dans la traduction anglaise. Dans ce dernier cas, c'est surtout le sémantisme du verbe en question ou de tout le contexte qui va véhiculer la notion de réfléchi. Sans être majoritaire, ce moyen est également utilisé dans la langue arabe. Un autre procédé est lui aussi commun aux deux langues, à savoir l'utilisation de *nafsun* et *self*.

2 Le réfléchi partiel

Donc, quoique la forme pronominale n'existe pas en tant que telle dans les langues arabe et anglaise, sa traduction demeure possible grâce aux moyens exposés plus haut. Cependant, un cas particulier intéresse notre étude, grâce à son originalité : il s'agit de ce que nous appelons le « réfléchi partiel ». Ce dernier se réfère à une situation particulière qui est celle relative à une partie du corps humain.

2.1 Traduction du réfléchi partiel

La raison pour laquelle le réfléchi partiel nous paraît particulièrement intéressant dans le cadre de cette étude, est qu'il remet en question les conclusions de la première partie de notre étude. En effet, le fait que toutes ces variantes sémantiques et morphosyntaxiques, apparaissant d'habitude dans la traduction des verbes pronominaux réfléchis dans le but de les différencier des verbes non pronominaux, disparaissent dans le cas du réfléchi partiel, nous paraît intéressant. L'exemple suivant concrétise parfaitement nos propos :

[11] Jean s'essuya les yeux.

| | | | |
|-------|----------------|------------|---------------------|
| [11'] | <i>masah-a</i> | <i>ḡān</i> | <i>'ajn-aj-hi.</i> |
| | « essuyer-Acc | Jean | œil-due acc – ses » |

[11''] *John wiped his eyes.* (qui veut littéralement dire : « Jean essuya ses yeux. »)

Comparons-le avec un autre exemple :

[12] Jean essuya son verre.

| | | | |
|-------|----------------|------------|-------------------|
| [12'] | <i>masah-a</i> | <i>ḡān</i> | <i>ka's-a-hu.</i> |
| | « essuyer-Acc | Jean | verre-acc-son » |

[12''] *John wiped his glass.*

La première remarque qui ressort de la comparaison des traductions de ces deux énoncés, est que la forme pronominale et la forme non pronominale sont traduites de la même façon à la fois dans les deux langues arabe et anglaise.

Concernant l'exemple 12, « essuya » est traduit par *masaha* et *wiped*, et « son verre » par *ka's-a-hu* et *his glass*. Ces deux traductions correspondent à la lettre aux données de la phrase d'origine. En effet, le possessif (*son* en français, *his* en anglais, et *-hu* en arabe) indique une relation d'appartenance entre le deuxième actant, complément de verbe (son verre) et le premier actant (Jean), (le verre appartient à Jean). Les deux langues arabe et anglaise ne se différencient pas de la langue française à ce niveau-là.

Par contre, dans le cas de l'exemple 11 « Jean s'essuya les yeux », les données des deux traductions arabe et anglaise ne coïncident plus avec celles de la phrase française, puisque les phrases arabe et anglaise traduisent la forme pronominale réfléchi respectivement par les adjectifs possessifs *-hu* et *his*.

Mais le possessif pourrait-il traduire toutes les notions contenues dans la structure pronominale ?

Il est vrai qu'une relation d'appartenance existe dans les deux cas (le réfléchi et le possessif). Mais c'est exclusivement sur cette relation d'appartenance que sont basées les traductions arabe et anglaise. En effet, dans la logique de ces deux langues, « Jean essuya les yeux de Jean » donne « Jean essuya ses yeux » tout comme « Jean essuya le verre de Jean » donne « Jean essuya son verre ». Là où il y a enchevêtrement dans les langues arabe et anglaise entre le possessif et le réfléchi, la langue française introduit une subtilité qui attribue une note spécifique à cette relation d'appartenance, dès qu'il s'agit de l'une des parties du corps humain. Ainsi, a-t-on dans la phrase française « il s'essuya [je précise quelle partie de son corps exactement] : les yeux ».

En effet, la langue française témoigne d'un « comportement particulier des parties du corps pour lesquelles la relation d'appartenance se marque par l'emploi, non d'un adjectif possessif, mais d'une construction pronominale. » (Gardes-Tamine 1998 : 164)

Donc, dans le cas du réfléchi partiel, la phrase pronominale française laisse supposer que l'être humain est son corps. Par contre, l'adjectif possessif des phrases arabe et anglaise pose le corps comme une possession de l'être humain au même titre que n'importe quel objet, en l'occurrence « le verre ».

Essayons de voir si d'autres données peuvent confirmer ce qui vient d'être énoncé. Dans ce but, considérons l'exemple suivant :

[13] Il s'est attribué le succès du projet.

[13'] *'azā nažāḥ-a l-mašrū'-i 'ila nafsih.*

[13"] *He gave himself credit for the success of the project.*

« Il s'est attribué » ou « * il a attribué à lui-même » est donc traduit en arabe par « *'azā 'ila nafsihī* », où *'azā* veut dire « attribuer », *'ila* est l'équivalent de « à », *nafs* veut dire « âme » et *hi* : « son ». Si nous revenons à l'étymologie du terme *nafs*, la traduction littérale de l'énoncé arabe sera : « *Il a attribué le succès du projet à son âme » ; même si cette réalité ne relève pas du domaine du conscient pour le locuteur arabe. Donc, dans la traduction de « lui-même », la langue arabe ne retient que l'âme. Dans cette langue, l'être humain est « réduit » à une âme, le corps n'est par conséquent plus considéré comme une composante intégrante de l'être humain.

Le corps et, par extrapolation, toute la vie sur terre ne sont en fin de compte, pour une personne croyante, qu'une étape de la vie éternelle. Il faudrait transcender ce corps pour atteindre l'immortalité. La langue arabe l'exprime très bien, en ne gardant de l'être humain que sa dimension spirituelle *nafsun* : « âme ».

En ce qui concerne la langue anglaise, la traduction de l'exemple précédent est similaire à la traduction arabe : *He gave himself credit for the success of the project*. Mais la grande différence avec l'énoncé arabe, c'est le mot *self* qui veut dire *own* ou *lui-même* en français, mais ne comporte aucun lien sémantique direct avec l'âme humaine comme c'est le cas du terme *nafs* de la langue arabe. C'est ainsi que le dictionnaire Merriam-Webster's, donne la définition suivante au terme *self* : « *The union of elements (as body, emotions, thoughts and sensations) that constitute the individuality and identity of a person.* »

2.2 Exemple à l'appui : « J'ai mal à la tête »

Jusqu'à cette étape, notre étude nous a montré que le rapport au corps n'est pas le même dans les trois langues. D'autres données peuvent être fournies à l'appui. Elles nous aideront à mieux préciser et nuancer notre point de vue.

[14] J'ai mal à la tête.

[14'] *ra's-ī ju-'lim-u-nī*.

[14"] *I have a headache.*

En effet, là où un locuteur français dit : « j'ai mal à la tête », un locuteur arabe déclare : « *ra's-ī ju-'lim-u-nī* ». Si on traduit littéralement cette formule, elle donnera en français « Ma tête me fait mal » qui, sans être agrammaticale, n'est pas pour autant la plus utilisée dans ce cas. Pour revenir à l'énoncé français étudié, nous dirons que « j'ai mal à la tête » montre l'identification entre *ma tête* et *moi*, le sens est « j'ai mal, [je précise où exactement] : la tête » ; *la tête* est donc partie intégrante du moi, alors que la formule arabe présente « ma tête » comme certes m'appartenant grâce à l'adjectif possessif *-ī*, mais qui reste toutefois extérieure à moi.

D'autre part, « ma tête me fait mal » (qui correspondrait exactement à la forme utilisée par la langue arabe) pourra être utilisée pour désigner un mal de tête

« extérieur », si je me suis cogné la tête contre le mur par exemple. Cette structure syntaxique devient en contrepartie très usitée en français dans des cas comme : « Son égoïsme me fait mal » (je ne pourrai certainement pas dire « *J'ai mal à son égoïsme ») :

[15] Son égoïsme me fait mal.

[15'] 'anānijjat-u-hu tu-'lim-u-nī.
 « égoïsme-nom.-son pers3 fém-faire mal-ind-me. »

[15''] *His selfishness hurts me.*

La structure syntaxique de l'énoncé arabe correspond exactement à celle utilisée pour exprimer le mal de tête : « ra's-ī- ju- 'lim-u- nī »
 « tête-ma pers3 masc-faire mal-ind-me ».

Ce qui révèle qu'un locuteur arabe s'exprime de la même façon en pensant « J'ai mal à la tête » et « Son égoïsme me fait mal ». « Son égoïsme » et « ma tête » sont placés au même niveau par rapport au locuteur.

Par contre, la langue anglaise traduit « j'ai mal à la tête » par « *I have a headache* » qui correspond mieux à la structure française. En introduisant *I = je*, la langue anglaise pose le mal de tête et par suite la tête comme partie intégrante du *je* et serait la traduction littérale de « j'ai un mal de tête ». La différence avec la structure syntaxique française est donc minime et ne touche pas l'essentiel.

Nous constatons que, dans ce cas, la langue anglaise est plus proche de la langue française quant à l'expression du rapport au corps.

En ce qui concerne l'énoncé « son égoïsme me fait mal », ses traductions anglaise « *His selfishness hurts me* » et arabe « 'anānijjat-u-hu tu-'lim-u-nī » lui correspondent à cent pour cent, mais il ne s'agit plus ici du rapport au corps.

Conclusion

On pourrait donc conclure que le rapport au corps n'est pas du tout identique dans les trois langues. Les langues française et arabe adoptent des attitudes extrêmes, la première en revendiquant la priorité du rapport au corps et la seconde en limitant l'être humain à sa dimension spirituelle. La langue anglaise, par contre, adopte une attitude médiane, en admettant le corps comme part constitutive de l'être humain, sans pour autant lui accorder un quelconque privilège.

La question de la relation au corps n'est d'ailleurs pas des plus simples, et cette étude ne prétend pas en avoir élucidé le mystère. La psychanalyse n'a pas encore épuisé le sujet, et Lou Andreas-Salomé dans sa *Lettre ouverte à Freud* a déclaré : « Notre propre corps n'est en effet rien d'autre que la part d'extériorité la plus proche de nous, inséparable de notre intimité, de notre identité ; mais nous en

sommes aussi coupés, au point qu'il nous faut apprendre à le connaître et à l'étudier de l'extérieur comme tout autre objet. Ainsi, dans nos relations d'objet, il est à la fois le champ de séparation qui nous coupe de tout le reste, et le lieu de rencontre de toute chose. » (1987 : 43)

Cette analyse pertinente de la part de l'amie et disciple de Sigmund Freud généralise la complexité de la question touchant ainsi l'être humain en général et non une culture en particulier. Toutefois, ce rapport au corps n'est pas vécu exactement de la même façon dans toutes les cultures. Il est vrai que l'être humain a toujours été considéré comme le lieu indiqué des luttes inlassables entre l'âme et le corps. Toutes les religions invitent l'être humain à dompter son corps pour atteindre l'immortalité. Le conflit n'est donc pas nouveau. Mais nous ne sommes pas sans savoir que Nietzsche avait déclaré en 1882 que, « Le plus grand événement récent – à savoir que « Dieu est mort », que la croyance au Dieu chrétien est tombée en discrédit commence dès maintenant à étendre son ombre sur l'Europe. » (1982 : 237). Si Dieu n'existe plus pour la société occidentale, l'âme à son tour n'existe plus. C'est alors le corps qui gagne du terrain dans l'identité et la vie de l'individu. En contrepartie, les sociétés orientales ont en général gardé plus longtemps cette dimension spirituelle et cette aspiration à l'immortalité à travers la transcendance du corps. C'est ce que les données linguistiques étudiées ont confirmé.

Il serait important de signaler à la fin que nous sommes conscients qu'une question aussi complexe que celle du rapport au corps mériterait des études plus amples, touchant d'autres données linguistiques des trois langues étudiées, et s'étendant à d'autres langues et par suite d'autres cultures. Mais son intérêt provient du fait que la traduction et la linguistique contrastive sont l'un des outils qui permettent d'éclairer certaines zones d'ombre, refoulées dans l'inconscient d'une expression linguistique collective, acquise et utilisée la plupart du temps automatiquement. Ce genre d'études s'avère utile pour une meilleure compréhension de soi et de l'autre, à un moment où il devient urgent que la logique du dialogue des cultures remplace celle du « choc des civilisations ».

Bibliographie

- Andreas-Salomé (L.), 1987 : *Lettre ouverte à Freud*, traduit par Dominique Miermont, Paris, Seuil (Collection Points).
- Blachère (R.) et Gaudefroy-Demombynes (M.), 1988 : *Grammaire de l'arabe classique*, Mayenne, Maisonneuve & Larose.
- Chomsky (N.), 1981 : *Réflexions sur le langage*, traduit par Judith Milner, Béatrice Vautherin, et Pierre Fiala. Paris, Flammarion.
- Gardes-Tamine (J.), 1998 : *La grammaire I*, 3^e édit., Paris, Armand Colin.
- Kouloughli (D.E.), 1994 : *Grammaire de l'arabe d'aujourd'hui*, Angleterre, Pocket (Langues pour tous).
- Maupassant (G.), 1982 : *Une vie*, France, LDP.
- Maupassant (G.), 1982 : *Siratu hajatin*, traduit par Elie Khalil. Beyrouth, Ouaydat.

- Maupassant (G.), 2009 : *A life*, traduit par : Roger Pearson, USA, Oxford University Press.
- Nietzsche (F.), 1982 : *Le gai savoir*, Paris, Gallimard.
- Riegel (M.), Pellat (J.-C.) et Rioul (R.), 1999 : *Grammaire méthodique du français*, 5^e édit., Paris, PUF.
- Saussure, (F.), 1998. *Cours de linguistique générale*. Paris, Payot & Rivages.
- Touratier (C.), 1996 : *Le système verbal français. Description morphologique et morphématique*, Paris, Armand Colin.

Zrinka Šimunić,
spes,
Zagreb,
Croatie,
simunic_zrinka@hotmail.com.

Analyse comparative du discours comme méthode d'évaluation de la qualité des productions textuelles lors d'une épreuve de traduction écrite

Résumé

L'analyse comparative du discours s'avère être une méthode efficace d'évaluation de la qualité des productions textuelles lors d'une épreuve de traduction écrite. D'une part, elle permet de répertorier les variations sémantiques et textuelles des discours traduits, d'établir le degré de leur écart par rapport au texte source et de déterminer leur poids respectif en fonction de leur rôle dans l'organisation du discours. D'autre part, elle contribue à une appréciation plus objective de l'impact des différents procédés et des différentes stratégies discursives sur la qualité des discours traduits en mettant en évidence l'ampleur des modifications de la structure sémantico-textuelle globale du texte source.

Mots-clés : discours traduits ; variations sémantiques ; variations textuelles.

« L'équivalence dans la différence est le problème cardinal du langage et le principal objet de la linguistique. » (Jakobson 1963 : 80)

Introduction

L'analyse des discours traduits suscite un intérêt grandissant chez les linguistes du discours, qui soulignent en premier lieu ses enjeux théoriques et méthodologiques. Comme le fait observer Jean-Michel Adam (en prenant l'exemple du discours politico-médiatique), la traduction en tant que phénomène de discursivité, c'est-à-dire abordée sous l'angle de l'analyse du discours, se révèle être au cœur du processus linguistique. L'analyse comparative des différents états du texte comme « construction produite en vue de la diffusion », qu'il s'agisse de ses « variations génétiques » ou de ses « variations par la traduction », permet de substituer à la recherche du « vrai texte » de tel ou tel grand discours « une conception du texte comme variation et somme différentielle de ses états successifs » et de situer ainsi la textualité entre « la constante stabilité acquise après quelques variations » et « la construction du sens comme source d'altérations » (Adam 2009).

La prise de conscience de l'importance des mécanismes textuels et sémantiques mobilisés au cours du processus de traduction et de leur part respective dans la structuration des discours traduits permet ensuite d'envisager un élargissement de l'analyse à différents types de discours traduits – des plus prestigieux aux plus ordinaires, mêmes si ces derniers sont moins bien structurés que les textes destinés à la publication et présentent un degré plus élevé de variation, voire d'écart par rapport au texte source. Ainsi, les discours traduits dans un contexte d'évaluation des apprentissages et des compétences soulèvent également la question de l'établissement du « vrai texte » et du statut de ses différentes variations, question d'autant plus délicate que les enjeux de leur analyse sont, avant tout, pédagogiques et professionnels.

La présente communication a pour objet ce dernier type de discours traduits. Elle vise à montrer comment leur analyse comparative, centrée sur quelques aspects bien précis de l'organisation du texte source et des textes cibles, peut contribuer à une évaluation plus objective de la qualité des productions textuelles lors d'une épreuve de traduction écrite. Son cadre théorique et méthodologique est constitué par le modèle genevois d'analyse du discours de type modulaire (Roulet 1999, 2000, 2001a, 2001b). Il s'agit d'un modèle d'analyse de la complexité de l'organisation du discours qui dispose d'un instrument d'analyse précis, susceptible d'être appliqué à l'étude systématique des différents types et genres du discours ou à une description plus approfondie des aspects isolés des productions discursives effectives, en fonction des objectifs fixés par l'analyste.

1 La présentation du corpus étudié

Le corpus étudié réunit 34 traductions d'un même texte source (une lettre d'affaires) réalisées dans un contexte de recrutement (un concours de recrutement de traducteurs organisé par une agence de traduction de Zagreb, Croatie) par des traducteurs professionnels, locuteurs natifs de la langue source (le croate) ou de la langue cible (le français). Dans l'analyse comparative des discours traduits, la problématique d'évaluation de la qualité des productions textuelles a été abordée sous l'angle de deux critères qui ont été communiqués aux candidats avant l'épreuve, à savoir la précision (l'exactitude, la fidélité) et le style (la clarté, l'élégance) de la traduction.

Un premier examen du corpus choisi a révélé des écarts considérables dans la manière de concilier ces deux exigences, qui n'ont pas été respectées dans la même mesure par les traducteurs natifs et non natifs de la langue cible, les premiers étant plus soucieux du style de la traduction, les seconds de sa fidélité au texte source. Ces écarts entre les discours traduits nous ont amenée à procéder d'abord à une analyse du texte source (TS) et du texte cible (TC) que nous avons sélectionné, intuitivement, après ce premier examen du corpus, comme texte cible répondant le mieux aux deux critères d'évaluation préétablis (TC₁). Les résultats de cette analyse préliminaire montrent que notre appréciation intuitive s'appuyait sur l'existence d'un haut degré d'équivalence sémantique entre le TS et le TC₁, voire sur

leur identité au niveau de la structure hiérarchique textuelle et de l'organisation relationnelle (Roulet 1999, 2000, 2001a, 2001b).

Le découpage en unités textuelles minimales du fragment discursif étudié (la séquence des actes [1] à [13]) et l'établissement de la structure hiérarchique textuelle et de l'organisation relationnelle du TS et du TC₁ serviront de point de départ pour une analyse comparative de l'ensemble des discours traduits, centrée sur leurs propriétés sémantiques et textuelles qui relèvent respectivement de deux critères d'évaluation préétablis.

[1] *TS Vezano uz Vašu prijavu za radno mjesto generalnog direktora,*

TC₁ Suite à votre candidature au poste de directeur général,

[2] *TS na žalost, moramo Vas obavijestiti da,*

TC₁ nous avons le regret de vous informer que

[3] *TS premda imate zadovoljavajuće kvalifikacije što se tiče Vašeg obrazovanja,*

TC₁ malgré vos qualifications satisfaisantes qui concernent en premier lieu votre niveau d'éducation,

[4] *TS nismo bili u mogućnosti uzeti Vas u užu izbor*

TC₁ nous n'avons pas pu vous présélectionner,

[5] *TS zbog nedostatka prakse iz navedenog područja.*

TC₁ une expérience pratique dans le domaine concerné vous faisant défaut.

[6] *TS Budući da ste ujedno predali i prijavu za radno mjesto voditelja prodaje,*

TC₁ Étant donné que vous avez également déposé un dossier pour le poste de chef des ventes,

[7] *TS za koje se ne traži takvo iskustvo,*

TC₁ qui ne prévoit pas une telle expérience,

[8] *TS smatramo da biste u toj ulozi bili dobrodošli u naš tim.*

TC₁ nous considérons que vous seriez, dans ce rôle, le bienvenu au sein de notre équipe.

[9] *TS Ukoliko ste za isto zainteresirani,*

TC₁ Si le poste de chef des ventes vous intéresse,

[10] *TS molimo Vas da to potvrdite našoj Kadrovskoj službi putem telefona, telefaksa ili e-maila,*

TC₁ nous vous prions de bien vouloir contacter notre service du personnel par téléphone, télécopie ou courrier électronique pour le confirmer,

[11] *TS kako bismo Vas mogli planirati za razgovor.*

TC₁ afin que nous puissions porter votre nom sur la liste des interviews.

[12] *TS Svoj pozitivni ili negativni odgovor trebate poslati najkasnije u roku 8 dana s naznakom « Natječaj ».*

TC₁ Une réponse positive ou négative est à envoyer dans un délai de 8 jours au plus tard, avec pour objet « Concours ».

[13] *TS U protivnom, smatrat ćemo da niste zainteresirani.*

TC₁ Dans le cas contraire, nous considérerons que vous n'êtes pas intéressé.

Le fragment discursif analysé forme une intervention textuelle complexe (*Figure 1*), dont la structure est caractéristique des échanges épistolaires (Roulet 1999 : 169-170). Elle est composée de trois interventions textuelles juxtaposées – deux interventions textuelles réactives et une intervention textuelle initiative.

La première intervention réactive correspond à la séquence des actes [1-5] et représente la réponse négative à la lettre de candidature pour le poste de directeur général mentionnée dans l'acte [1] (*prijava*: « candidature »). La deuxième intervention réactive correspond à la séquence des actes [6-8] et représente la réponse positive à la lettre de candidature pour le poste de chef des ventes mentionnée dans l'acte [6] (*prijava*: « dossier »). L'intervention initiative, qui correspond à la séquence des actes [9-13], est une demande de confirmation de candidature pour le deuxième poste, évoquée dans les actes [10] (*potvrditi*: « confirmer ») et [12] (*odgovor*: « réponse »).

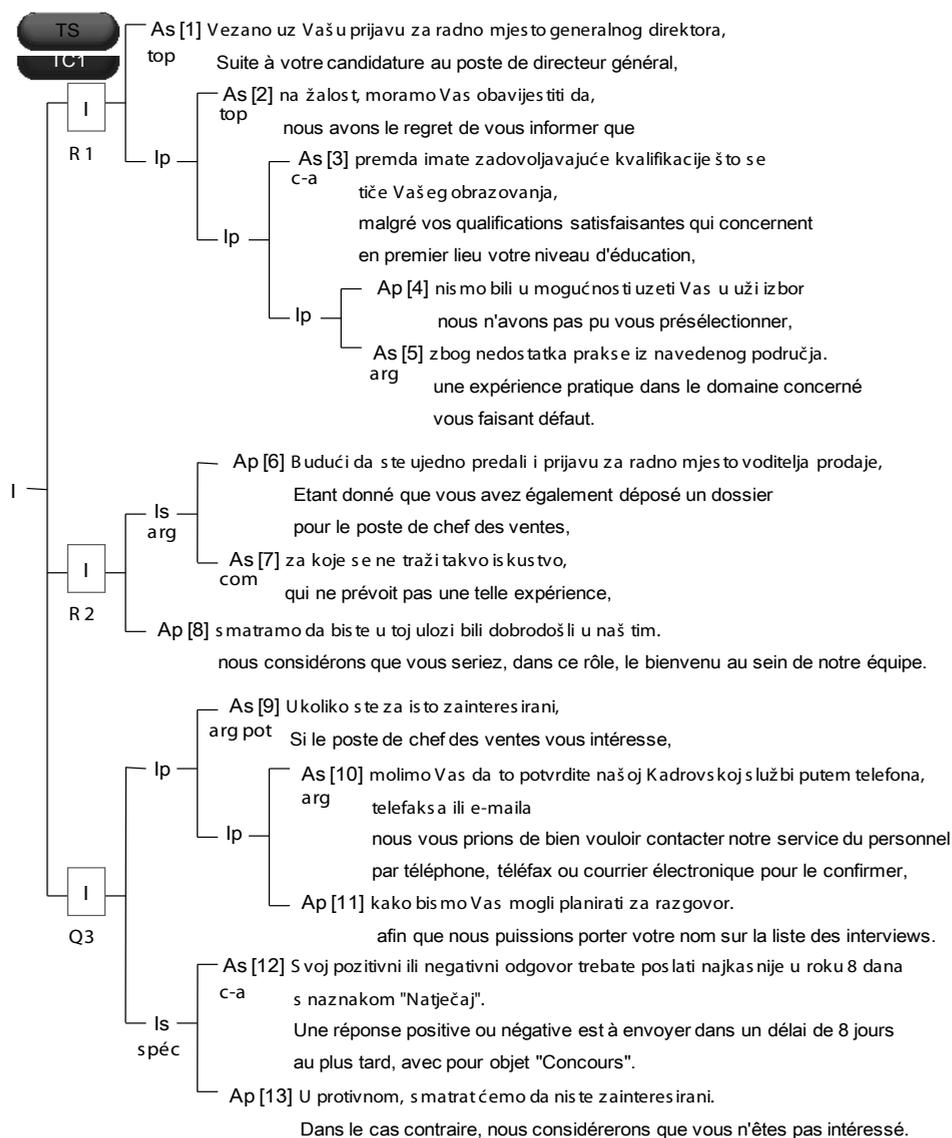


Figure 1 : l'établissement de la structure hiérarchique textuelle et de l'organisation relationnelle du TS et du TC1

Il s'agit d'une structure textuelle dynamique, qui demande au traducteur la maîtrise des procédés discursifs spécifiques à la communication commerciale (Whittaker 2001), liés notamment au ménagement de la face du destinataire (Goffman 1967, Brown & Levinson 1987, Kerbrat-Orecchioni 1992). En outre, le fragment discursif analysé présente une organisation sémantique et relationnelle particulièrement dense, ce qui oblige le traducteur à recourir à des stratégies discursives plus ou moins élaborées face aux pièges et difficultés liés à sa traduction.

2 Variations sémantiques des discours traduits

L'analyse sémantique devrait nous permettre de déterminer le degré d'équivalence sémantique entre les actes textuels qui composent les discours traduits et les unités textuelles minimales correspondantes du TS. Une attention particulière sera accordée aux actes textuels qui, d'une traduction à l'autre, présentent le plus haut degré de divergences et qui, de ce fait, semblent avoir posé aux traducteurs le plus grand nombre de difficultés d'ordres lexical ou syntaxique.

Les variations dans la traduction de l'acte [3] sont liées à la présence du mot polysémique *kvalifikacije* (« qualifications »). *Le Petit Robert* distingue trois sens principaux de ce nom en français, qui sont identiques en croate. Nous retiendrons la troisième définition qui se rapporte aux « qualifications professionnelles » et la distinction entre « ensemble des aptitudes et des connaissances acquises d'un travailleur pour l'exercice d'une activité de production » et « formation à une fonction technique spécialisée », distinction qui s'appuie sur la représentation conceptuelle suivante :

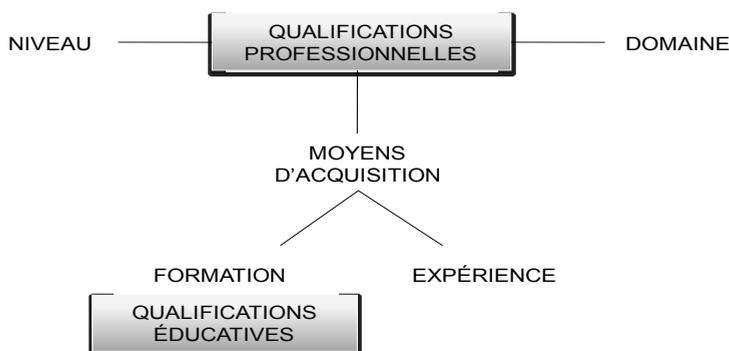


Figure 2

Dans le texte source, le mot *qualifications* a été employé dans son sens plus étroit de « formation », voire de « qualifications éducatives ». Une éventuelle incertitude liée au sens du mot *qualifications* est écartée dans les actes [4] et [5], où « le manque d'expérience pratique » est évoqué comme motif de refus de la candidature.

Cependant, dans les discours traduits que nous avons analysés, le nom *qualifications* a été utilisé aussi bien au sens large de « qualifications professionnelles » qu'au sens plus étroit de « formation », le choix de l'une ou de l'autre interprétation influant sur la structure textuelle de cette séquence d'actes. En outre, plusieurs traducteurs ont substitué à l'adjectif *satisfaisant*, qui est assez ambigu dans les deux langues (« acceptable », « bon », « convenable », mais aussi « assez bon », « passable », « moyen ») un adjectif ou une locution verbale plus valorisants pour le destinataire.

L'examen de notre corpus permet de distinguer différents degrés de variation sémantique des discours traduits par rapport au texte source. Dans un premier groupe de discours traduits, la traduction de l'acte [3] se rapproche le plus de son sens dans le texte source :

- [3] TC7 en dépit du fait que votre formation soit satisfaisante
 TC8 bien que votre formation corresponde tout à fait à ce poste
 TC10 malgré votre formation appropriée

Un deuxième groupe est constitué par les discours traduits où le mot *qualifications* (ou son équivalent) se rapporte à l'ensemble des qualifications professionnelles, ce qui a entraîné l'ajout d'un nouvel acte textuel relié au premier par une relation interactive de spécification (marquée par *notamment*, *principalement*, etc.) :

- [3a, 3b] TC5 bien que vous possédiez les qualifications nécessaires, et nous nous référons ici principalement à votre formation
 TC15 bien que vous possédiez les qualifications nécessaires pour cette fonction, et notamment en termes de formation
 TC3 en dépit de vos qualités notamment de votre éducation

Un troisième groupe rassemble les discours traduits qui sont sémantiquement plus riches que le texte source. On peut y observer la surdétermination du mot *qualifications* en vue d'une désambiguïsation rétroactive de l'acte [3] (traduit comme un seul acte ou comme deux actes distincts) par l'ajout du mot *niveau*, par opposition au mot *domaine* dans l'acte [5] – comme le montre la structure conceptuelle « émergente » (Filliettaz 2001) de ces deux actes :

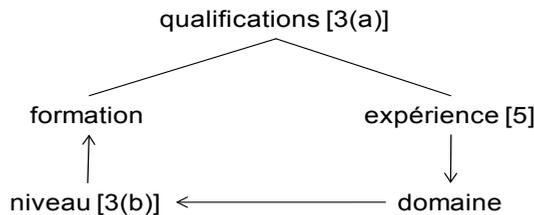


Figure 3

- [3] TC1 malgré vos qualifications satisfaisantes qui concernent en premier lieu votre niveau d'éducation
 [3a, 3b] TC11 malgré vos bonnes qualifications professionnelles, et plus particulièrement votre niveau de formation

À la différence de l'acte [3], dont l'ambiguïté sémantique est à l'origine des variations dans sa traduction, l'acte [4] permet d'illustrer les variations sémantiques des discours traduits liées à la présence d'un vocabulaire précis qui n'est pas maîtrisé dans la même mesure par tous les traducteurs. Il s'agit, en l'occurrence, de la locution verbale *uzeti u uži izbor* (« retenir sur la liste restreinte de candidats », « retenir en sélection finale »), qui désigne une étape précise dans le déroulement du concours.

Nous pouvons distinguer différents degrés de précision des discours traduits par rapport au texte source, sans tenir compte des traductions mot-à-mot, inappropriées dans la langue cible :

[4] TC16 nous n'avons pas été en mesure de vous prendre dans notre choix serré

Les exemples suivants illustrent le cas des traductions appropriées, qui ont gardé le même degré de précision que le texte source :

[4b] TC33 il ne nous a été pas possible de vous retenir en sélection finale

TC1 nous n'avons pas pu vous présélectionner

Certaines traductions, que l'on peut qualifier de partielles, se sont limitées à la première partie de la locution verbale (*retenir*) :

[4c] TC12 nous ne sommes pas en mesure de retenir votre candidature

Cependant, la plupart des traducteurs ont substitué à la locution initiale un verbe ou une locution verbale désignant une étape dans le déroulement du concours différente de celle indiquée dans le texte source. Il s'agit le plus souvent de l'étape initiale (*prendre en considération*) ou finale (*donner une suite favorable, répondre positivement à la demande*) :

[4d] TC25 nous n'avons pas pu vous prendre en considération

TC34 nous n'avons pas été dans la possibilité de vous recruter

TC9 nous sommes au regret de vous informer que nous ne pouvons pas donner une suite favorable à votre demande

TC23 malheureusement nous sommes dans l'impossibilité de répondre positivement à votre demande pour le poste désigné ci-dessus

À noter que certains de ces choix sont non seulement plus imprécis que le texte source, mais aussi plus menaçants pour la face du destinataire (par exemple (*ne pas*) *prendre en considération*).

Dans l'acte [5], les variations dans la traduction sont liées aux différentes possibilités d'exprimer la causalité. Comme ces variations relèvent en même temps de l'organisation relationnelle des discours traduits, elles seront analysées de manière détaillée dans la section suivante.

3 Variations textuelles des discours traduits

L'analyse textuelle devrait nous permettre d'évaluer la manière dont les unités discursives minimales sont intégrées dans les structures textuelles et relationnelles des discours traduits et le degré d'identité de celles-ci avec les structures des TS et TC1.

La complexité de la structure hiérarchique textuelle et de l'organisation relationnelle du fragment discursif étudié – liée à la combinaison dans la même lettre de

deux visées communicationnelles opposées (le refus de candidature pour le premier poste et l'acceptation de candidature pour le deuxième poste) – se manifeste par la fréquence des marqueurs de relations textuelles. Nous avons soumis à un examen systématique l'emploi de ces marqueurs afin d'établir le degré de leurs divergences d'un discours traduit à l'autre, la nature de ces divergences et leurs effets sur la qualité des productions textuelles (Berrendonner 1983, Rossari 1993, 2000, Roulet 2006).

Nous avons pu observer que certains de ces marqueurs ont fait l'objet de traductions très variées. Ainsi, le connecteur argumentatif *z bog* dans l'acte [5] et le connecteur contre-argumentatif *u protivnom* dans l'acte [13] ont donné lieu respectivement à 17, voire 12 traductions différentes, des traductions appropriées assurant l'équivalence textuelle entre le TS et le TC aux traductions littérales ou inexactes, sources de leurs divergences sémantico-textuelles (voir en *Annexe*). L'analyse des variations liées à l'emploi des marqueurs de relations textuelles a montré que chez les locuteurs natifs de la langue cible, ces variations sont principalement d'ordre stylistique et visent l'amélioration de la qualité de la traduction, tandis que chez les locuteurs non natifs de la langue cible, elles sont davantage liées à leurs compétences discursives inégales qui peuvent parfois donner lieu à des traductions inadéquates, littérales ou trop linéaires de certains de ces marqueurs et, par conséquent, ont des effets négatifs sur la qualité de leurs productions textuelles.

L'analyse comparative de la structure textuelle et de l'organisation relationnelle des unités discursives plus vastes dans lesquelles s'inscrivent ces deux actes textuels nous a permis d'observer, tout d'abord, une relative stabilité textuelle des discours traduits, en dépit de (ou, peut-être, grâce à) ces nombreuses variations dans la traduction liées au choix des marqueurs de relation textuelle argumentative dans l'acte [5] et contre-argumentative dans l'acte [13]. La structure textuelle de la plupart des discours traduits est identique à celle du TS.

Dans les discours traduits dont la structure textuelle n'est pas identique à celle du TS, nous avons pu observer trois types de variations textuelles et leurs différentes combinaisons : modifications du nombre (extension, compression ou suppression) d'actes textuels, modification de leur ordre et modification des relations interactives entre les constituants textuels de rangs différents.

Ces variations textuelles se manifestent à des degrés différents d'un discours traduit à l'autre – de celles qui se limitent aux constituants subordonnés de la structure hiérarchique textuelle à celles qui affectent les relations interactives entre ses principaux constituants.

La traduction de la première intervention dans le TC₁₁ offre l'exemple de modifications limitées aux constituants subordonnés de la structure textuelle du TS (*Figure 4*).

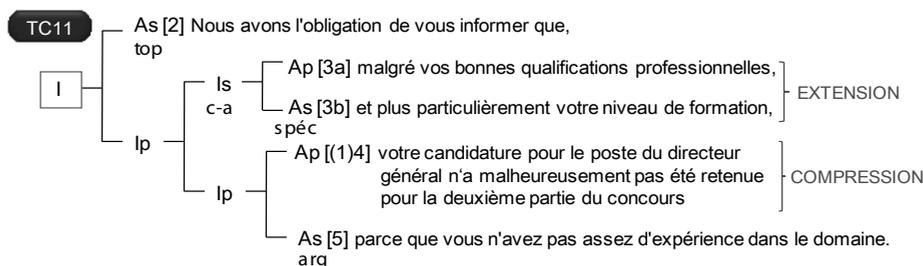


Figure 4 : Modifications limitées aux constituants subordonnés de la structure textuelle

La traduction de la même intervention dans le TC10 a donné lieu aux trois types de modifications de la structure textuelle et de l'organisation relationnelle du TS : la réduction du nombre d'actes textuels, la modification de leur ordre (source d'ambiguïté du discours traduit) et la modification des relations interactives entre les principaux constituants de la structure textuelle du TS (Figure 5).

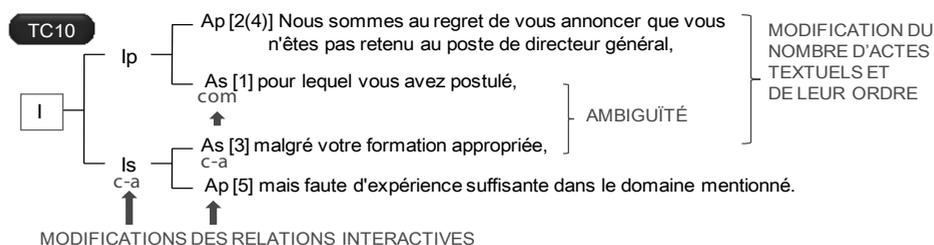


Figure 5 : Modifications qui affectent les relations entre les principaux constituants de la structure textuelle

La traduction de la deuxième intervention dans le TC23 offre un autre exemple de modifications qui affectent les relations entre les principaux constituants de la structure textuelle du TS (Figure 6). En outre, l'insertion d'un connecteur qui n'est pas marqué dans le TS (*en revanche*) entraîne la modification des relations interactives au niveau de l'unité textuelle plus vaste.

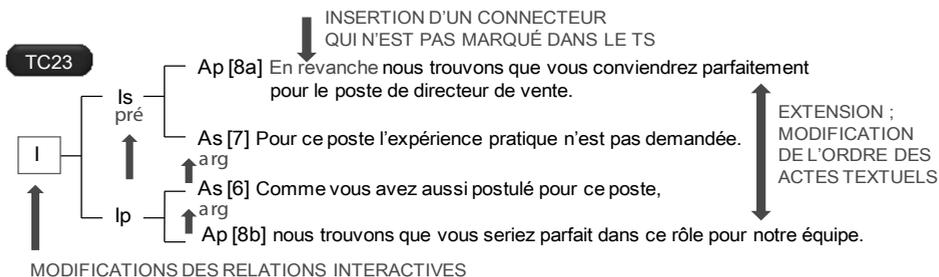


Figure 6 : Insertion d'un connecteur qui n'est pas marqué dans le TS

Finalement, la traduction de la troisième intervention dans le TC6 offre l'exemple de variations textuelles qui affectent les relations interactives entre les principaux constituants de la structure du TS suite à la modification de la portée du connecteur. Dans le TS, le connecteur *dans le cas contraire* lie l'acte [13] à l'acte [12] par une relation interactive contre-argumentative, alors que, dans le TC6, il enchaîne sur le connecteur *dans le cas où* inséré par le traducteur dans l'acte [9] (Figure 7).

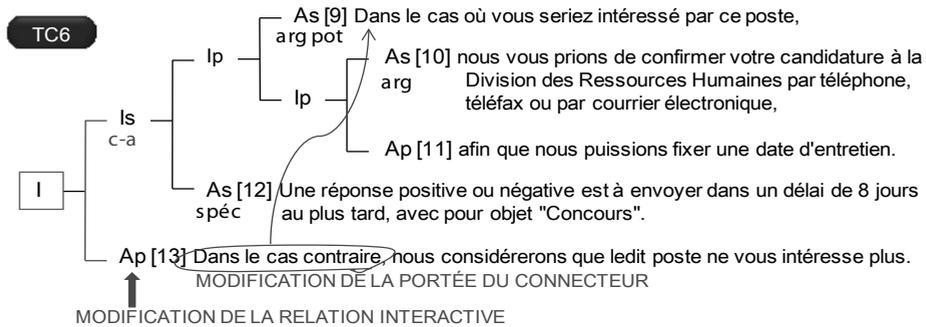


Figure 7 : Modification (élargissement) de la portée du connecteur

Les variations textuelles que nous avons répertoriées peuvent avoir un impact plus ou moins important sur la qualité des productions textuelles, qu'il s'agisse de variations d'ordre stylistique, qui visent l'amélioration de la qualité des discours produits ou de variations qui font partie des stratégies discursives déployées par le traducteur face aux difficultés rencontrées. Ce dernier type de variations, susceptible d'entraîner des écarts sémantico-textuels considérables des discours traduits par rapport au TS, mérite d'être approfondi.

4 Principales stratégies discursives déployées dans la structuration sémantico-textuelle des discours traduits

La mise en relation des aspects sémantiques et textuels de l'organisation des discours traduits fait ressortir deux principales stratégies discursives déployées par les traducteurs face aux difficultés rencontrées : la stratégie de reformulation et la stratégie d'évitement. Ces deux stratégies discursives résultent de différentes combinaisons de variations sémantiques et textuelles dans les discours traduits. Nous les aborderons ici sous l'angle de leur impact respectif sur la qualité des discours traduits.

La stratégie de reformulation combine les variations sémantiques visant la clarification du contenu sémantique de certaines unités lexicales ou syntaxiques, avec différents types de variations textuelles. Son impact sur la qualité des productions textuelles est proportionnel à l'étendue de ces variations. Dans l'exemple du TC11 (Figure 4), il est négligeable : d'une part, l'extension textuelle de l'acte [3] est limitée aux constituants subordonnés de la structure textuelle et, d'autre part,

la compression des actes [1] et [4] n'entraîne pas la modification de leur contenu sémantique. Dans l'exemple du TC₂₃ (Figure 6), l'impact de la stratégie de reformulation est plus grand – l'extension de l'acte [8] entraîne la modification des relations interactives entre les principaux constituants de la structure textuelle du TS.

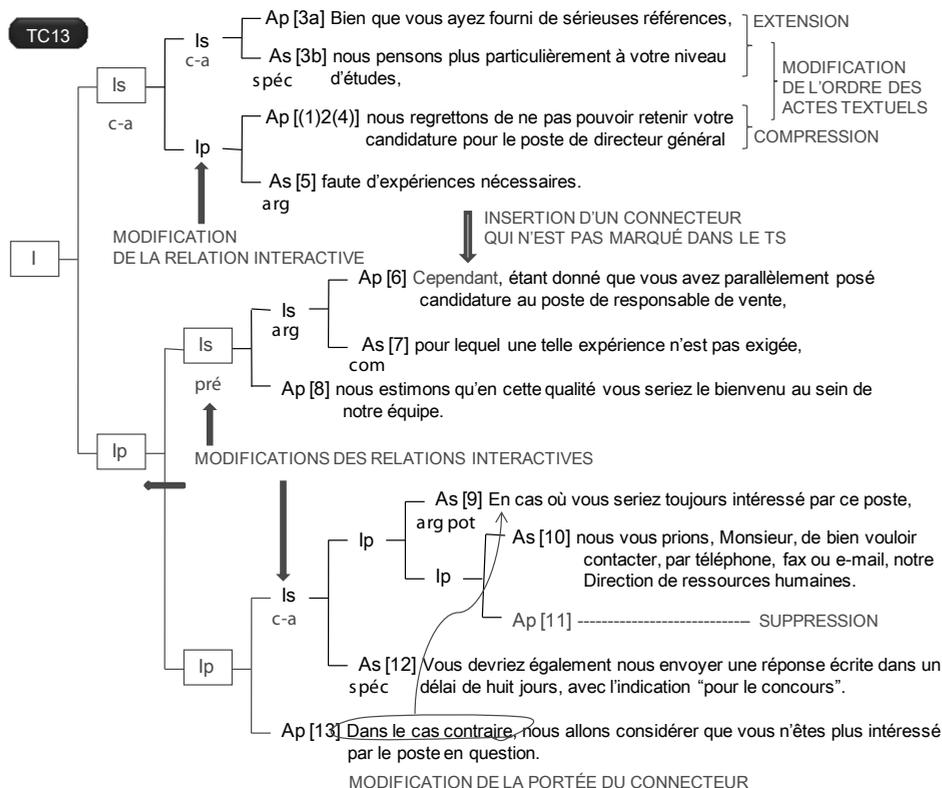


Figure 8 : Accumulation de plusieurs types de modifications dans le TS entraînant une transformation qualitative de la structure sémantico-textuelle du TS

En recourant à la stratégie d'évitement, le traducteur modifie la structure du TC en supprimant les éléments qui lui semblent problématiques dans le TS. Ce genre de procédé peut entraîner non seulement des modifications sémantiques du texte source mais aussi des remaniements importants de sa structure textuelle. Ainsi, dans l'exemple du TC₁₀ (Figure 5), la compression des actes [2] et [4], liée à la suppression partielle du contenu sémantique de l'acte [4] dans le TS, entraîne une modification complète de l'organisation relationnelle de la première intervention.

Dans les segments textuels plus étendus, on peut observer (dans certains discours traduits) la combinaison de plusieurs procédés qui relèvent de l'un ou de l'autre type de stratégie discursive, ou des deux à la fois. Leur accumulation dans le même discours traduit peut entraîner une transformation qualitative de la structure sémantico-textuelle du TS, comme nous le montre l'exemple du TC₁₃ (Figure 8).

Conclusion

Les discours traduits dans un contexte d'évaluation des apprentissages et des compétences contiennent de nombreuses variations sémantiques et textuelles par rapport au texte source. Ces variations peuvent être fondées sur divers motifs et avoir divers impacts sur la structure sémantico-textuelle globale des discours traduits.

Après avoir rappelé l'enjeu théorique de l'analyse comparative des propriétés sémantiques et textuelles des discours traduits et sa contribution à la redéfinition des frontières de la textualité, nous avons essayé de mettre en évidence ses enjeux pratiques en étudiant les possibilités de son utilisation comme méthode d'évaluation de la qualité des productions textuelles lors d'une épreuve de traduction écrite. L'étude d'un corpus de textes authentiques nous a amenée à conclure que l'analyse comparative de l'organisation sémantico-textuelle du texte source et des textes cibles peut contribuer de plusieurs manières à une évaluation plus objective de la qualité des discours traduits.

Premièrement, elle permet de répertorier les différentes variations des discours traduits par rapport au texte source. Nous avons pu constater que la présence des mots polysémiques, l'emploi d'un vocabulaire précis et la fréquence des marqueurs de relations textuelles sont à l'origine de la plupart des variations sémantiques répertoriées, tandis que les variations textuelles sont liées à la modification du nombre d'actes textuels, à la modification de leur ordre et à la modification des relations interactives entre les constituants textuels de rangs différents.

Deuxièmement, un examen plus approfondi de ces variations permet d'établir le degré de leur écart par rapport au texte source et de déterminer leur poids respectif en fonction de leur rôle plus ou moins important dans l'organisation du discours. Ainsi les variations sémantiques, si elles ne touchent pas à la structure textuelle, ont une portée moins importante que les modifications du sens du texte source qui s'étendent à des segments textuels plus longs. De même, les variations textuelles qui n'affectent pas la structure sémantique ont une portée moins importante que les variations textuelles qui dissimulent une modification du sens du texte source.

Troisièmement, l'analyse comparative des discours traduits permet de relier leurs aspects sémantiques et textuels et d'évaluer l'impact des différents procédés et stratégies discursives sur la qualité des productions textuelles en mettant en évidence l'ampleur des modifications de la structure sémantico-textuelle globale du texte source. Les différentes combinaisons des variations sémantiques et textuelles sur lesquelles s'appuient deux principales stratégies discursives déployées par les traducteurs, la stratégie de reformulation et la stratégie d'évitement, nous laissent entrevoir les intentions qui les motivent (désambiguïser le texte source, faciliter la lecture en langue cible, résoudre une difficulté de traduction liée à l'absence d'équivalence entre deux langues, contourner un problème faute de temps/compétences/ressources linguistiques nécessaires, etc.) et nous amènent à distinguer les modifications justifiées

de celles résultant de l'insertion des actes discursifs qui sont porteurs d'intentions propres aux traducteurs et qui constituent des traces de leur subjectivité.

Bibliographie

- Adam (J.-M.), 2005 : *La linguistique textuelle: Introduction à l'analyse textuelle des discours*, Paris, Armand Colin.
- Adam (J.-M.), 2009 : « Construction des données textuelles en analyse de discours : établissement du texte et/ou prise en charge de ses variations ? », conférence plénière au III^e Colloque international interdisciplinaire *Le français parlé dans les médias*, 1-4 septembre 2009, Lausanne, www.unil.ch/fpm2009/page59037.html.
- Berrendonner (A.), 1983 : « « Connecteurs pragmatiques » et anaphore », dans *Cahiers de linguistique française*, n° 5, p. 215-246.
- Brown (P.) et Levinson (S.), 1987 : *Politeness. Some universals in language usage*, Cambridge, Cambridge University Press.
- Filliettaz (L.), 2001 : « La dimension référentielle », dans Roulet (E.), Filliettaz (L.), Grobet (A.) et Burger (M.), *Un modèle et un instrument d'analyse de l'organisation du discours*, Berne, Peter Lang, p. 97-137.
- Goffman (E.), 1967 : *Interaction ritual. Essays on face-to-face behavior*, New York, Anchor Books.
- Jakobson (R.), 1963 : « Aspects linguistiques de la traduction », dans *Essais de linguistique générale: 1. Les fondations du langage*, Paris, Les Éditions de Minuit, p. 78-86.
- Kerbrat-Orecchioni (C.), 1992 : *Les interactions verbales* (vol. II), Paris, Armand Colin.
- Rossari (C.), 1993 : *Les opérations de reformulation*, Berne, Lang.
- Rossari (C.), 2000 : *Connecteurs et relations de discours: des liens entre cognition et signification*, Nancy, Presses Universitaires de Nancy.
- Roulet (E.), 1999 : *La description de l'organisation du discours: du dialogue au texte*, Paris, Didier.
- Roulet (E.), 2000 : « Une approche modulaire de la complexité de l'organisation du discours », dans Adam (J.-M.) et Nølke (H.), éd., *Approches modulaires: de la langue au discours*, Lausanne, Delachaux & Niestlé, p. 187-257.
- Roulet (E.), 2001a : « La dimension hiérarchique », dans Roulet (E.), Filliettaz (L.), Grobet (A.) et Burger (M.), *Un modèle et un instrument d'analyse de l'organisation du discours*, Berne, Peter Lang, p. 53-95.
- Roulet (E.), 2001b : « L'organisation relationnelle », dans Roulet (E.), Filliettaz (L.), Grobet (A.) et Burger (M.), *Un modèle et un instrument d'analyse de l'organisation du discours*, Berne, Peter Lang, p. 165-199.
- Roulet (E.), 2006 : « The description of text relation markers in the Geneva model of discourse organization » (Chapter 7), dans Kerstin (F.), éd., *Approaches to Discourse Particles*, Amsterdam, Elsevier, p. 115-131.
- Šimunić (Z.), 2006 : « Modularni pristup kompleksnosti organizacije diskursa », dans *Rasprave Instituta za hrvatski jezik i jezikoslovlje*, n° 31, février 2006, p. 277-328.
- Whittaker (S.), 2001 : « La correspondance commerciale: apprentissage de stratégies discursives en langue étrangère », dans *Revue française de linguistique appliquée* 2001/2, Vol. VI, p. 95-102.

Annexe: Marqueurs de relations textuelles dans le texte source (TS) et les textes cibles (TC1-TC34)

| TS | vezano uz | premda | zbog | ∅ | budući da | ukoliko | kako | u protivnom |
|------|--------------------|-----------------|---------------|-------------|------------------------------|----------------|------------|--------------------------|
| TC1 | suite à | malgré | -ant | ∅ | étant donné que | si | afin que | dans le cas contraire |
| TC2 | suite à | bien que | par suite de | ∅ | comme | si | afin que | sans |
| TC3 | ∅ | en dépit de | du fait de | ∅ | étant donné que | si | afin que | si cela n'est pas le cas |
| TC4 | suite à | malgré | en raison de | ∅ | du fait que | si | afin de | dans le cas contraire |
| TC5 | au sujet de | bien que | vu que | ∅ | étant donné que | au cas où | pour | dans le cas contraire |
| TC6 | suite à | malgré | do à | ∅ | étant donné que | dans le cas où | afin que | dans le cas contraire |
| TC7 | à la suite de | en dépit de | pour | ∅ | étant donné que | si | afin de | dans le cas contraire |
| TC8 | ∅ | bien que | car | ∅ | vu que | si | pour | dans le cas contraire |
| TC9 | en nous référant à | bien que | pour autant | ∅ | étant donné que | si | afin que | dans le cas contraire |
| TC10 | ∅ | malgré | mais faite de | toutesfois | et au vu que | si | afin de | sans |
| TC11 | ∅ | malgré | parce que | cependant | ∅ | si | après quoi | faute de quoi |
| TC12 | concernant | malgré | à cause de | ∅ | étant donné que | dans le cas où | pour que | dans le cas contraire |
| TC13 | ∅ | bien que | faute de | cependant | étant donné que | en cas où | ∅ | dans le cas contraire |
| TC14 | suite à | malgré | à cause de | ∅ | puisque | dans le cas où | afin que | dans le cas contraire |
| TC15 | concernant | bien que | car | ∅ | étant donné que | si | afin de | autrement |
| TC16 | concernant | même si | à cause de | ∅ | en tenant compte du fait que | si | afin de | au contraire |
| TC17 | faisant suite à | malgré | étant donné | ∅ | étant donné le fait que | si | afin que | au contraire |
| TC18 | en ce qui concerne | bien que | à cause de | ∅ | vu que | si | pour que | au contraire |
| TC19 | suite à | malgré | à cause de | ∅ | ∅ | si | pour que | si cela n'est pas le cas |
| TC20 | suite à | malgré | faute de | ∅ | prenant en compte que | si | afin de | en cas contraire |
| TC21 | concernant | bien que | vu | ∅ | comme | si | pour | sinon |
| TC22 | au sujet de | même que | à cause de | ∅ | prenant en compte que | si | pour que | au cas contraire |
| TC23 | concernant | en effet...mais | parce que | en revanche | comme | si | pour que | dans le cas contraire |
| TC24 | ∅ | malgré | ∅ | ∅ | puisque | si | pour que | dans le cas contraire |
| TC25 | en nous référant à | malgré | à cause de | ∅ | vu que | si | pour que | contrairement |
| TC26 | concernant | quoique | car | ∅ | puisque | si | pour que | sinon |
| TC27 | concernant | même si | à cause de | ∅ | vu que | si | pour que | sinon |
| TC28 | au sujet de | quoique | suite à | ∅ | vu que | si | pour que | au contraire |
| TC29 | concernant | bien que | faute de | ∅ | étant donné que | si | pour | au contraire |
| TC30 | à propos de | malgré | parce que | ∅ | puisque | si | pour | au contraire |
| TC31 | concernant | même que | à cause de | ∅ | puisque | si | pour que | contrairement |
| TC32 | au sujet de | même que | à cause de | ∅ | vu que | si | pour que | sinon |
| TC33 | en réponse à | bien que | en raison de | ∅ | étant donné que | si | pour que | dans le cas inverse |
| TC34 | concernant | même si | à cause de | ∅ | ayant en compte | si | ainsi | dans le cas contraire |

Kaja Gostkowska,
Institut de philologie romane,
Université de Wrocław,
Wrocław,
Pologne,
kaja.gostkowska@wp.eu.

Les terminologies française et polonaise du génie biomédical face au défi de la diversité

Résumé

Dans notre article, nous visons à décrire comment deux langues différentes, le français et le polonais, se comportent face au développement rapide des sciences. Nous présentons cette problématique en nous basant sur l'exemple du domaine du génie biomédical (et plus précisément, de la biomécanique) et de sa terminologie dans ces deux langues. Nos considérations se focalisent autour de deux grands axes : d'abord, nous examinons la question des emprunts externes pour, ensuite, passer à l'étude des termes transférés entre les disciplines et des manifestations linguistiques du changement de point de vue accompagnant ce processus.

Mots-clés : terminologie ; interdisciplinarité ; génie biomédical ; biomécanique ; emprunt ; point de vue.

Introduction

Les débuts de la recherche dans le domaine de la biomécanique à l'École polytechnique de Wrocław datent des années 1970. À cette époque, les ingénieurs du laboratoire de mécanique se sont mis à explorer le fonctionnement du corps humain à l'aide des lois de la mécanique en commençant une coopération avec les médecins (surtout avec les chirurgiens). Rapidement, ils se sont rendu compte que leur coopération ne serait pas facile. La cause en était très simple : ils n'utilisaient pas le même discours spécialisé. Il a donc fallu un certain temps pour que les spécialistes des différents domaines puissent élaborer un discours commun et qu'ils puissent adopter une nouvelle terminologie provenant de ce croisement des disciplines. La terminologie biomécanique est donc relativement nouvelle et, en plus, elle se développe dans un contexte interdisciplinaire par excellence. Ces deux facteurs constituent un véritable défi, non seulement pour les spécialistes qui sans cesse introduisent des termes nouveaux, mais aussi pour les terminologues qui devraient contribuer à l'harmonisation de cette terminologie.

Dans notre communication, nous visons à présenter comment deux langues différentes, à savoir le français et le polonais, se comportent face à ce défi. Nous nous demandons également quels sont les moyens lexicaux et grammaticaux qui permettent à chaque langue de constituer sa propre terminologie dans un contexte professionnel dominé fortement par la langue anglo-américaine. Nous avons restreint le domaine étudié à la biomécanique qui fait partie d'un domaine beaucoup plus vaste, le génie biomédical. Notre analyse est basée sur un corpus de termes et de contextes tirés de textes spécialisés français et polonais. Il a été constitué à partir de monographies ainsi que de recueils d'articles dont les auteurs représentent des disciplines et des milieux professionnels différents : parmi eux se trouvent des médecins, mais aussi des ingénieurs de différentes spécialités comme des mécaniciens ou des électroniciens.

Dans un premier temps, nous voulons attirer l'attention sur le statut différent du français et du polonais qui se répercute au niveau de la politique linguistique. Nous allons ensuite présenter brièvement la biomécanique et son champ de recherche pour pouvoir ensuite relever les indices de sa diversité au niveau terminologique. Nous nous intéresserons en particulier aux moyens morphologiques utilisés pour créer des termes nouveaux, ainsi qu'au phénomène de transfert des termes entre les disciplines au sein de la même langue.

1 Statut respectif des deux langues

Dans notre travail, nous visons à comparer deux langues différentes non seulement du point de vue purement linguistique (donc morphologique, syntaxique, lexical, etc.), mais aussi au niveau extralinguistique. On sait bien que toutes les langues n'ont pas la même valeur politique ou sociale. Dans le modèle gravitationnel proposé par Louis-Jean Calvet (1999 : 76-78), qui présente une typologie des langues du monde en fonction de leur prestige et de leur importance en tant que langues de communication, le français joue un rôle de langue dite « supercentrale », et le polonais se trouve parmi les langues centrales. Au centre de ce système se place l'anglais en tant que langue hypercentrale. En outre, les deux langues étudiées appartiennent à deux communautés culturelles et scientifiques différentes. Le français représente une communauté qui a créé des théories scientifiques très importantes pour le développement technologique et intellectuel de l'humanité, tandis que le polonais est utilisé par une communauté linguistique qui n'a pas de tels acquis au niveau européen ou mondial. La tradition scientifique dans les deux pays varie donc considérablement. Tous ces facteurs se reflètent au niveau de la politique linguistique. Les activités de normalisation en terminologie sont beaucoup mieux développées et, surtout, plus réputées dans les communautés francophones. En Pologne, certes, elles existent aussi, mais ne connaissent pas une ampleur ni une reconnaissance comparables à celles qu'on observe en France.

2 Présentation de la discipline étudiée

Avant de passer à la description des deux terminologies, il est indispensable de présenter brièvement la discipline étudiée. La biomécanique dont il sera question ici fait partie d'un champ de recherche plus large, à savoir celui du génie biomédical. Le génie biomédical essaie de synthétiser des connaissances très diverses sur le corps humain et sur son fonctionnement en vue d'améliorer la qualité de vie des hommes. Son appellation anglaise *Bioengineering* n'a été utilisée pour la première fois qu'en 1954 par un chercheur anglo-allemand, Heinz Wolff. Depuis, son champ de recherche est devenu de plus en plus large et il englobe toujours de plus en plus de disciplines, l'une d'elles étant la biomécanique.

Celle-ci focalise ses recherches sur le fonctionnement moteur du corps pour concevoir les dispositifs artificiels (prothèses, implants, etc.) pouvant remplacer les éléments défaillants de l'appareil moteur humain. Selon les auteurs d'un ouvrage publié dans la collection *Que sais-je ?* (Allard et Blanchi, 1999 : 3), la biomécanique est une étude « des structures de tous les organismes vivants et de leurs fonctions par rapport à l'environnement [...] faisant appel aux lois de la physique (optique, mécanique, électrique, etc.) ». Elle est née au carrefour de domaines comme la physique, la médecine, l'informatique ou la chimie, qui constituent ses véritables disciplines mères et auxquelles elle emprunte de la terminologie et de la méthodologie. C'est donc un champ de recherche interdisciplinaire par excellence qui, en plus, se construit sous nos yeux.

3 Le défi résultant du développement rapide des sciences

La terminologie biomécanique est, comme nous venons de le souligner, en permanente création. En plus, elle est utilisée dans un milieu qui est fortement marqué par la présence de l'anglo-américain, qualifié par François Gaudin de « véritable *lingua franca* des sciences contemporaines » (2003 : 198). Il en résulte que cette terminologie possède un nombre important d'emprunts externes, donc des emprunts « faits à une autre langue » (Bessé 1992 : 186). Une définition plus élaborée dit que l'emprunt est « la matrice qui fait introduire dans des énoncés français des lexies existant dans d'autres langues et absentes dans un état immédiatement antérieur de la langue française » (Sablayrolles et Jacquet-Pfau 2008 : 21). Les emprunts relevés dans nos corpus français et polonais seront présentés en deux sous-classes : d'abord les emprunts lexicaux et, ensuite, les emprunts structuraux (également appelés *calques*).

3.1 Les emprunts lexicaux

Les emprunts lexicaux sont des mots « intégralement transférés (la forme et le sens du mot sont empruntés) » (Bessé 1992 : 25). Dans la terminologie biomédicale, ils constituent un groupe assez important. Il convient de signaler que certaines formes, empruntées majoritairement à l'anglais, se rencontrent en français et en polonais, par exemple les termes *stress-shielding*, *remodeling* ou *stent*. Il faut ajouter que les termes français ont déjà fait l'objet d'une certaine adaptation à la langue cible dans la

mesure où ils ne sont plus écrits entre guillemets et sont précédés d'un article français. L'adaptation peut aller plus loin, de sorte que des morphèmes grammaticaux propres à la langue cible remplacent ceux d'origine étrangère. Dans la terminologie française, l'anglicisme *remodeling* possède son équivalent francisé *remodelage*.

La terminologie polonaise, quant à elle, fournit des exemples de termes qui ne sont pas encore adaptés. Ils sont cités entre guillemets sous leur forme anglaise et, en plus, accompagnés d'un mot générique polonais, par ex. : *mechanizm* « *weeping lubrication* » ou *koncepcja* « *low friction arthroplasty* ». Pour ce qui est de ce deuxième terme, il est fort probable qu'il ne sera pas polonisé, parce que le concept auquel il renvoie semble déjà dépassé. En outre, l'emploi de la forme anglo-américaine, surtout dans la première phase d'adaptation d'un terme, ne devrait pas être blâmé, puisque cela permet d'éviter certains malentendus. Dans la terminologie polonaise, on peut rencontrer deux termes anglo-américains référant à la méthode chirurgicale d'implantation des prothèses, à savoir *press-fit* et *exact-fit*. Ils apparaissent avec des équivalents polonais très divers : le terme *press-fit* peut accompagner les termes polonais *wzmężone ciśnienie*, *podwyższone ciśnienie* ou encore *ściste dopasowanie*. Le terme *exact-fit*, quant à lui, accompagne soit le terme *czynnościowe dopasowanie* soit le terme *ściste dopasowanie*. On peut espérer qu'une telle abondance de termes pouvant conduire à des quiproquos trouvera sa solution soit dans la norme officielle adoptée à ce sujet, soit dans la plus grande fréquence d'emploi d'un de ces termes, ou bien encore dans la réalité extralinguistique qui va naturellement éliminer ces notions (ce qui semble être d'ailleurs le cas, puisque ces méthodes chirurgicales se sont avérées inefficaces).

3.1.1 Les internationalismes

Un certain nombre d'emprunts, vu leur spécificité, sera traité à part. Il s'agit des internationalismes, c'est-à-dire des mots qui apparaissent sous la même forme dans au moins trois langues appartenant à des familles différentes. Les termes qui remplissent cette condition peuvent être divisés selon différents critères. D'abord, on distingue des termes qui comportent soit un seul lexème (figure 1) :

| anglais | polonais | français | allemand | russe | portugais |
|------------------|----------------|-----------------|------------------|------------------|-----------------|
| <i>prothesis</i> | <i>proteza</i> | <i>prothèse</i> | <i>Prothese</i> | (био-)протез | <i>prótese</i> |
| <i>implant</i> | <i>implant</i> | <i>implant</i> | <i>Implantat</i> | <i>имплантат</i> | <i>implante</i> |

Figure 1

soit deux lexèmes (figure 2) :

| anglais | polonais | français | allemand | russe | portugais |
|-------------------------------|-------------------------------|-------------------------|--------------------------------|---------------------------------|-----------------------------|
| <i>biomedical engineering</i> | <i>inżynieria biomedyczna</i> | <i>génie biomédical</i> | <i>biomedizinische Technik</i> | <i>биомедицинская инженерия</i> | <i>Engenharia Biomédica</i> |

Figure 2

Plusieurs internationalismes sont construits à l'aide de formants savants, par exemple de formants grecs comme *bio-*, *ostéo-*, *endo-*, *arthro-*, *allo-* ou *télé-* (figure 3):

| anglais | polonais | français | allemand | russe | portugais |
|-------------------------|------------------------|-------------------------|-------------------------|------------------------|------------------------|
| <i>biomechanics</i> | <i>biomechanika</i> | <i>biomécanique</i> | <i>Biomechanik</i> | <i>биомеханика</i> | <i>biomecânica</i> |
| <i>biomaterials</i> | <i>biomateriały</i> | <i>biomatériaux</i> | <i>Biomaterialien</i> | <i>биоматериалы</i> | <i>biomaterial</i> |
| <i>osseointegration</i> | <i>osteointegracja</i> | <i>ostéointégration</i> | <i>Osseointegration</i> | <i>остеоинтеграция</i> | <i>Osseointegração</i> |
| <i>endoprosthesis</i> | <i>endoproteza</i> | <i>endoprothèse</i> | <i>Endoprothese</i> | <i>эндопротез</i> | <i>endoprótese</i> |
| <i>arthroplasty</i> | <i>artroplastyka</i> | <i>arthroplastie</i> | <i>Arthroplastik</i> | <i>артропластика</i> | <i>artroplastia</i> |
| <i>alloplasty</i> | <i>aloplastyka</i> | <i>aloplastie</i> | <i>Alloplastik</i> | <i>аллопластика</i> | <i>aloplastia</i> |
| <i>telemedicine</i> | <i>telemedycyna</i> | <i>télé médecine</i> | <i>Telemedizin</i> | <i>телемедицина</i> | <i>Telemedicina</i> |

Figure 3

La question qui se pose est de savoir quelle serait la véritable langue source de ces termes. Notre étude n'est certes pas la première (ni la dernière) tentative d'entamer une telle réflexion. Comme le font remarquer Jean-François Sablayrolles et Christine Jacquet-Pfau (2008 : 24) : « Nombre d'innovations se répandent rapidement et mondialement, avec les termes qui les dénomment, sans qu'il soit possible d'étudier leur trajectoire ». Plus loin, les auteurs admettent : « Néanmoins, l'avancée technologique des USA fait que ces mots sont souvent d'origine anglo-américaine » (2008 : 24). Nous ne pouvons que souscrire à cette constatation.

Paradoxalement, ce procédé de formation des termes semble la preuve d'une certaine liberté ou d'une certaine indépendance de chaque langue et de ses usagers. Sans pouvoir l'expliquer, on constate que les spécialistes français préfèrent le terme français *ancrage biologique* à son équivalent internationalisé *ostéointégration*; de même en polonais, on parlera de *biogodność* d'un matériau plutôt que de sa *biokompatybilność* (*biocompatibilité* en français s'inscrit dans la série des internationalismes).

3.2 Les emprunts structuraux (calques)

Bruno de Bessé (1992 : 181) définit les calques comme la « transposition d'un élément d'une langue dans une autre, par traduction ». Leur fréquence d'apparition est assez élevée, aussi bien en français qu'en polonais. La portée de ce procédé néologique ne devrait pas être sous-estimée, surtout lorsqu'on prend en compte le fait que des emprunts structuraux identiques existent simultanément en français et en polonais. Le terme anglo-américain *functional bone adaptation* est traduit comme

adaptation fonctionnelle et *funkcjonalna adaptacja kości*. De même, le terme *wear products* apparaît en français sous la forme de *débris d'usure* et en polonais comme *produkty zużycia*. Plusieurs autres termes polonais constituent une bonne illustration des calques : *polymer-tissue construct* est traduit en *konstrukt polimerowo-tkankowy*, *multimodal interface* en *interfejs multimodalny* et *minimally invasive surgery* en *chirurgia minimalnie inwazyjna*. Le nombre considérable de calques parmi tous les nouveaux termes biomédicaux semble confirmer le statut de l'anglo-américain en tant que langue internationale de la science.

4 Le défi de l'interdisciplinarité

La science contemporaine est fortement marquée par le phénomène de l'interdisciplinarité. Les recherches sont de plus en plus détaillées, les spécialisations se multiplient, les disciplines traditionnelles voient sans cesse se renouveler leurs champs de recherche. Comme le constate Jean-Paul Resweber, « tout se passe comme si, à l'intérieur du cloisonnement traditionnel des disciplines, le domaine livré à la recherche se révélait curieusement mobile, extrêmement complexe et étrangement discontinu » (1981 : 33). Dans le cadre de la biomécanique, il s'agit d'une coopération et d'un dialogue des spécialistes représentant des disciplines très diverses. De leurs confrontations résulte une synthèse de différents points de vue adoptés à propos d'un même objet, ce qui constitue la particularité de l'approche interdisciplinaire. Sur le plan linguistique, les termes sont transférés entre les disciplines. Comme le souligne Pascaline Dury (1999 : 17-18) : « de nombreuses sciences, pour ne pas dire toutes [...], empruntent des concepts à d'autres sciences, connexes ou non, et sont le centre d'une véritable circulation langagière des termes ». Sur le plan conceptuel, on fait appel à des concepts déjà existants, ce qui permet de « puiser dans un réseau conceptuel organisé, et fonctionnant déjà dans une autre discipline » (Dury 1999 : 24).

La question qui se pose est de savoir si ce transfert des termes est accompagné d'un changement sur le plan sémantique et conceptuel dans le domaine en question. Suite à l'analyse de notre corpus nous avons pu constater que le sens de certains termes empruntés n'a pas changé, que leurs valeurs conceptuelles sont restées les mêmes. La biomécanique se sert par exemple de termes physiques ou mathématiques pour décrire des réalités nouvelles sans pour autant créer de nouvelles acceptions de ces termes. La situation est beaucoup plus complexe en ce qui concerne les termes empruntés à l'anatomie ou à la physiologie. Au premier abord, leur sens ne change pas, mais en même temps, ils apparaissent dans des contextes tout à fait différents. Une étude plus approfondie nous amène, ensuite, à la constatation faite par Pascaline Dury : « quelle que soit l'ampleur de la circulation d'un terme, et le nombre de transgressions sémantiques qu'il subit, il subsiste toujours un sens noyau, qui indique son origine » (1999 : 22). Le sens noyau des termes empruntés à l'anatomie reste stable mais, en même temps, le fait d'être utilisés dans des contextes nouveaux leur procure de « nouveaux caractères sémantiques » (Dury 1999 : 21).

Dans le cadre de la biomécanique, les termes empruntés à l'anatomie renvoient à de mêmes objets mais perçus d'une autre manière. Ce changement de perspective peut s'observer dans l'entourage le plus proche du terme. Nous allons présenter brièvement quelques manifestations linguistiques parmi les plus frappantes de ce changement de perspective. Nous ne présentons ici que des exemples tirés du corpus français en signalant que les mêmes phénomènes apparaissent dans les textes polonais.

4.1 Les expressions du point de vue

Dans les textes spécialisés, on trouve tout un inventaire d'expressions qui servent à marquer dans le texte, *expressis verbis*, le point de vue adopté. Leur inventaire n'est pas très large, mais ils apparaissent fréquemment. Les spécialistes ne laissent pas de place aux doutes, mais soulignent l'appartenance de leur raisonnement et de leurs méthodes de recherche à une discipline précise :

- [1] Cette première articulation métatarso-phalangienne ne fonctionne donc pas comme une rotule **au sens mécanique du terme**. (Bo : 258)
- [2] **Du point de vue biomécanique**, c'est le péroné qui est l'os le plus important des deux os de la jambe [...]. (Bo : 333)
- [3] Sa [le cartilage] constitution tripartite à base de collagène, de protéoglycans et de chondrocytes en fait un tissu hautement spécialisé **dans l'ordre mécanique**. (Bo : 469)
- [4] Pauwels a beaucoup contribué à donner à l'arthrose sa véritable signification **sur le plan biomécanique**. (Bo : 474)
- [5] [...] ainsi très rapidement, la tête fémorale s'emboîte **de façon mécaniquement satisfaisante**. (Hc : 32)

4.2 Le changement de point de vue inscrit dans les définitions

Avant de passer à la présentation des exemples, il faut souligner qu'il ne s'agit en aucun cas de définitions strictement terminologiques, qui pourraient être utilisées dans les dictionnaires. Les spécialistes essaient d'établir une définition plutôt simplifiée, ils visent à déterminer les caractéristiques essentielles d'un objet et de lui attribuer un sens nouveau selon la perspective adoptée :

- [6] **L'os est un matériau composite** formé de collagène et d'hydroxyapatite. (Qsj : 60)
- [7] Il faut se rappeler que **le corps humain est un système mécanique polyarticulé commandé** qui est composé de chaînes cinématiques arborescentes ouvertes (support unipodal) ou fermée (support bipodal) et doté d'un nombre élevé de degrés de liberté. (Qsj : 98)

4.3 Les contextes spécialisés

Le contexte le plus proche du terme peut également comporter des indices de changement de point de vue. Dans la discipline étudiée, les termes d'anatomie ont

été tirés de leur domaine d'origine et se sont ancrés dans des contextes nouveaux : ils sont accompagnés de verbes et adjectifs différents, ils font partie de nouvelles locutions. Cet environnement linguistique est la meilleure preuve d'une nouvelle perspective, du fait qu'on envisage les objets qui se cachent derrière ces termes d'un autre point de vue. Ainsi :

- [8] **Les os ont un comportement mécanique** similaire à ceux des métaux. (Qsj : 59)
- [9] **Les caractéristiques mécaniques du tissu osseux** varient suivant l'âge, le sexe, le type d'os, et la région de l'os lui-même [...]. **L'os est** donc beaucoup moins **résistant en torsion et en compression** radiale qu'en traction et compression longitudinale. [...] **L'os offre une plus grande résistance mécanique aux efforts rapides qu'aux efforts lents** [...]. (Bo : 34)
- [10] [...] si **le tibia n'a qu'un rôle de transmission des contraintes** en flexion et extension, c'est le péroné qui intervient essentiellement **dans la transmission des forces** en rotation : mouvement dans lequel l'os est le plus fragile. (Bo : 333)
- [11] Le cartilage reposant sur son socle osseux **a pour but d'assurer le glissement ainsi que la transmission, la répartition et l'amortissement des contraintes**. (Bo : 469)

Les exemples ci-dessus montrent que les entités anatomiques sont envisagées d'un point de vue complètement différent, qu'elles sont considérées comme des composants d'un système mécanique sur lequel agissent différentes forces et contraintes régies par les lois de la physique. Les contextes fournissent des informations sur les propriétés physiques et sur les fonctions des entités anatomiques. Comme le remarque Hélène Beciri, ces informations ne seraient pas retenues comme pertinentes dans le cadre d'une définition, mais elles « facilitent la tâche du lecteur en lui permettant de se faire une représentation concrète de l'objet désigné » (2003 : 49).

4.4 Les comparaisons

La dernière catégorie linguistique relevée, qui manifeste le changement du point de vue, comporte des comparaisons. Dans ce type de structures, les entités anatomiques sont décrites en tant que systèmes mécaniques très particuliers :

- [12] Selon le modèle de Hill, **le muscle fonctionne comme un système mécanique** constitué par une composante élastique en série avec une composante contractile. (Qsj : 51)
- [13] [...] **le cartilage** [...] protège les articulations contre les forces d'impact **en agissant comme un amortisseur**. (Qsj : 61)

Cette catégorie semble très productive. Les spécialistes disposent de moyens linguistiques très nombreux pour créer des images suffisamment évocatrices, même pour les non-spécialistes. Leur choix ne se limite pas à l'utilisation de *comme* mais ils ont recours à d'autres moyens lexicaux et syntaxiques :

- [14] L'amortissement des pressions et des vibrations est à notre avis fondamental : sans cette propriété, **notre appareil locomoteur serait comparable à une voiture sans pneu ni suspension**. (Bo : 479)
- [15] Ce système en équilibre peut être assimilé à celui d'une balance dont le fémur serait la colonne portante et le bassin le fléau. (Hc : 55)
- [16] [...] les forces induites par les **muscles** périarticulaires qui **sont de véritables serre-joints articulaires** [...]. (Bo : 331)

Les comparaisons concernent principalement la construction et le fonctionnement des éléments anatomiques. Elles établissent une sorte d'équivalence entre deux entités appartenant à deux domaines différents.

- [17] **Le couple mécanique os-muscle** qui forme une entité mécanique va se comporter **comme une poutre composite**. (Bo : 2)
- [18] Les muscles n'agissent donc pas sur les leviers squelettiques **à la façon des câbles d'une grue**, mais réalisent avec eux un système solidaire permettant de résister aux efforts qui leur sont imposés. (Bo : 334)

La fonction principale de ce type de structures comparatives semble être de rendre des entités anatomiques plus familières du point de vue conceptuel aux spécialistes d'un autre domaine. Ainsi, les spécialistes de la biomécanique, partant de la perspective de leur discipline d'origine, conçoivent des entités nouvelles, essaient de comprendre leur fonctionnement et leur structure.

Conclusion

L'analyse de la terminologie biomécanique française et polonaise nous permet d'observer quelques phénomènes importants liés au développement de la science d'aujourd'hui. D'un côté, on assiste à l'apparition de champs de recherche interdisciplinaire qui naissent au carrefour des disciplines traditionnelles comme la médecine, la biologie, la chimie ou la physique. De l'autre côté, depuis déjà plusieurs années, on constate l'apparition de nouveaux termes à un rythme très accéléré. Nous avons essayé de montrer la façon dont ces facteurs se reflètent dans la terminologie biomécanique dans deux langues différentes. La comparaison du français et du polonais montre des différences surtout au niveau de la création des termes nouveaux par le biais des emprunts. La communauté francophone, grâce à une tradition normalisatrice plus longue et importante, présente une plus forte tendance à adapter les termes empruntés et à créer leurs équivalents français. En ce qui concerne le polonais, l'action des mécanismes normalisateurs, dont ceux qui portent sur des disciplines en train de naître, est beaucoup plus faible.

L'analyse de la terminologie biomécanique permet également d'étudier le phénomène de transfert des termes et des concepts d'une discipline scientifique à l'autre. Grâce à l'étude des termes et de leurs contextes, nous avons présenté quelques manifestations

linguistiques du changement de point de vue qui accompagne ce processus. Nous sommes consciente cependant du fait que cette problématique, si caractéristique de la science contemporaine, exige des études encore plus approfondies. L'important est de tenir compte du fait que l'étude de la terminologie en situation interdisciplinaire nécessite une approche spécifique. Comme le souligne Pascaline Dury, ces vocabulaires spécialisés « ne sont pas hermétiques les uns par rapport aux autres, mais tissent des liens sémantiques qui les font envisager comme des territoires aux frontières perméables, plutôt que comme des univers clos » (1999 : 25).

Bibliographie

Ouvrages théoriques

Beciri (H.), 2003 : « Néologismes et définition en contexte : pour une typologie des indices interprétatifs formels », dans Steuckardt (A.) et Niklas-Salminen (A.), éd., *Le mot et sa glose*, Publications de l'Université de Provence, p. 41-56.

Bessé (B. de), 1992 : *Cours de terminologie. Notes de cours*, Université de Genève.

Calvet (L.-J.), 1999 : *Pour une écologie des langues du monde*, Plon.

Dury (P.), 1999 : « Les variations sémantiques en terminologie : étude diachronique et comparative appliquée à l'écologie » dans Delavigne (V.) et Bouveret (M.), éd., *Sémantique des termes spécialisés*, Publications de l'Université de Rouen, p. 17-32.

Gaudin (F.), 2003 : *Socioterminologie. Une approche sociolinguistique de la terminologie*, Bruxelles, Éditions Duculot.

Resweber (J.-P.), 1981 : *La méthode interdisciplinaire*, Paris, Presses Universitaires de France.

Sablayrolles (J. F.) et Jacquet-Pfau (Ch.), 2008 : « Les emprunts : du repérage aux analyses. Diversité des objectifs et des traitements », dans *Neologica. Revue internationale de la néologie*, n° 2, 2008, p. 19-38.

Sources d'exemples

Amh : Allard (P.) et Blanchi (J.-P.), 2000 : *Analyse du mouvement humain par la biomécanique*, Décarie, Éditions Vigot.

Bf : Dufour (M.) et Pillu (M.), 2005 : *Biomécanique fonctionnelle. Membres – Tête – Tronc*, Paris, Masson.

Bo : *Biomécanique orthopédique*, Poitout (D. G.), dir., 1987, Paris, Masson.

Bpm : Bouisset (S.), 2002 : *Biomécanique et physiologie du mouvement*, Paris, Masson.

Hc : Castaing (J.), 1979 : *La hanche*, Anatomie fonctionnelle de l'appareil locomoteur 4., Paris, Éditions Vigot.

Qsj : Allard (P.) et Blanchi (J.-P.), 1999 : *La biomécanique*, Que sais-je?, Paris, Presses Universitaires de France.

Dardo de Vecchi,
Euromed-Management, Marseille,
Université Paris Diderot – EILA,
France,
dardo.devecchi@euromed-management.com,
devecchi@eila.univ-paris-diderot.fr.

Les mots des entreprises : des termes mal connus

Résumé

Les entreprises et les organisations façonnent la langue générale et la langue spécialisée dans leur secteur d'activité pour énoncer leur réalité quotidienne en produisant un matériau linguistique propre : le parler d'entreprise. Ce parler se caractérise par des unités linguistiques qui montrent les concepts utilisés, ainsi, leur nature est terminologique. La pragmatérminologie les aborde en rapport avec les acteurs et la culture de la communauté qui les a façonnés. Ainsi, ils ont une valeur ajoutée pour le traducteur et pour l'institution elle-même qui trouvent en la langue une matière première à exploiter autant en situation monolingue que plurilingue.

Mots-clés : entreprise ; organisation ; parler d'entreprise ; terme maison ; pragmatérminologie.

Introduction

L'observation de la langue en entreprise n'est pas une tâche facile ni courante. Si une partie de la langue utilisée est visible de l'extérieur (sites internet, publicités, rapports, communiqués, etc.), la plus grande partie est à l'intérieur de l'organisation et n'est pas accessible aux observateurs extérieurs. Il n'y a en effet pas de raison pour que l'entreprise donne libre accès à son univers interne, surtout si ce n'est pas elle-même qui le sollicite. Aussi la bibliographie est-elle limitée, la confidentialité de rigueur. Lorsqu'il s'agit d'étudier les mots¹ des entreprises, la vue que l'on peut avoir de l'extérieur n'est que partielle : ce que l'on voit est la partie émergée d'une masse considérable d'un matériau linguistique dont on ne connaît pas les implications à l'intérieur. De plus, en entreprise tout n'est pas écrit et beaucoup d'énoncés restent oraux : tout ce qui se dit n'est pas susceptible d'être écrit ou destiné à l'être. Et pourtant, l'oral donne beaucoup d'informations à propos des pratiques des entreprises. Les mots observables de l'extérieur sont donc le résultat d'un processus qui commence à l'intérieur.

¹ Pour les besoins de la présentation, nous utiliserons les conventions typographiques suivantes : des chevrons pour les concepts : < concept > ; des guillemets pour les termes : « terme » (en dehors des citations).

Aujourd'hui, le linguiste observateur et le gestionnaire locuteur sont devant le fait accompli : l'entreprise ne se dit pas dans une langue, mais dans des langues, diversité, globalisation et localisation obligent. L'enjeu de la traduction est là et par la même occasion celui du lexicologue qui étudie les mots de la langue et du terminologue qui étudie les termes d'un domaine. La question qui se pose alors est la suivante : que faut-il exactement traduire : des mots, des termes ? Nous dirions : ni les uns ni les autres, dans la mesure où ce qui est nécessaire *in fine*, c'est que l'entreprise puisse occuper une place dans un marché parlant une autre langue. Ne l'oublions pas, publiques ou privées, les entreprises sont des unités de production de biens et de services et pour avoir une place dans un marché, il convient qu'elles se distinguent des autres. Les langues en entreprise jouent alors un rôle essentiel et c'est dans les expressions utilisées par ses membres que l'on peut observer la langue en action.

Les acteurs en entreprise connaissent leurs expressions, les reconnaissent et s'y reconnaissent : c'est leur culture propre. Nous postulons que ce ne sont pas les mots des langues en entreprise qui ont une valeur particulière, mais que ce sont les termes énoncés dans les langues spécialisées pour l'exploitation d'une activité qui l'ont. L'analyse lexicologique risque de ne dévoiler qu'une partie du phénomène observé, tandis que la terminologie offre d'autres perspectives pour le traitement des langues en entreprise. Ce que les acteurs conçoivent puis expriment – par quelque système sémiotique que ce soit – ce sont des termes. Ils créent un réseau solidaire qui s'insère dans les besoins mêmes des entreprises pour exister dans un marché. L'analyse terminologique de ce réseau nécessite à notre avis une approche particulière : celle de la pragmatérminologie qui s'adresse directement aux organisations, dont les entreprises sont un cas particulier. L'entreprise se dit en langue et en le faisant elle produit sa propre terminologie ; sans elle l'entreprise ne peut exister.

Lucrèce écrivait à propos de l'origine du langage : *At varios linguae sonitus natura subegit/mittere, et utilitas expressit nomina rerum* (« Quant aux divers sons du langage, c'est la nature qui poussa les hommes à les émettre, et c'est le besoin qui fait naître les noms des choses ») (1971 : 88). On remarquera la subtilité de l'opposition/complémentarité entre *natura* et *utilitas*. Toute distance gardée, la situation s'applique aux entreprises et aux organisations d'aujourd'hui. Leurs termes existent parce que l'entreprise en a besoin, parce qu'ils sont l'aboutissement d'un processus de conceptualisation, voire de la réflexion d'un groupe de travail ; les acteurs ont besoin de repères communs. L'utilité étant le moteur de la présence de ces termes, ils sont le produit de connaissances ancrées dans des pratiques culturelles, d'où leur valeur identitaire. C'est bien de cet ensemble que l'approche pragmatérminologique a l'ambition de rendre compte.

Après une brève réflexion à propos des besoins de traduire, nous étudierons quelques exemples destinés à montrer l'écart entre les places respectives des mots et des termes dans l'entreprise. Ensuite, nous exposerons ce que l'approche pragmatérminologique permet de mettre en relief.

1 L'entreprise traduit pour des publics et pour des pays

Les besoins d'exportation ou d'implantation à l'étranger ont mis depuis longtemps les entreprises, quels que soient leur taille et leur métier (en gestion, on parle du « métier » de l'entreprise pour faire référence à son activité), devant le besoin de traduction. La globalisation, les voyages et surtout l'explosion d'internet ont augmenté ce besoin de manière exponentielle. En outre, en dehors du cadre de la globalisation, dans un pays multilingue, nonobstant le statut de ses langues, il est très approprié de s'adresser à ses clients dans leur langue. Par exemple en Espagne, la coopérative Eroski a très bien compris l'intérêt de traduire son site dans quatre langues d'Espagne : basque, castillan, catalan et galicien. Elle traduit aussi dans ces quatre langues les étiquettes de sa propre marque de distribution. En Suisse, le supermarché Migros fait de même : le site est proposé en allemand, français et italien. Il s'agit de s'adresser sur son propre territoire à des publics nationaux appartenant à des origines linguistiques différentes. Ce sont des marchés qu'on doit considérer.

L'entreprise peut aussi être amenée à proposer des traductions pour des publics non nationaux sur son propre territoire. Si l'on opère dans le secteur du tourisme, la traduction s'impose : treize langues pour le site de l'Office du tourisme de Paris, six pour un opérateur privé comme Cityrama dans la même ville. À Londres, le *British Museum* affiche huit langues. En France, les grandes écoles font de même, cinq langues pour l'ESCP Europe (ex-École supérieure de commerce de Paris), trois pour Euromed-Management.

Lorsque les entreprises s'implantent à l'étranger, la traduction est fondamentale. Le magasin d'électroménager français Darty est présent en Italie, au Luxembourg et en Turquie. À chaque fois Darty traduit son site dans les langues dans lesquelles il est nécessaire de réaliser son activité. Et, pour le Luxembourg, le site fait apparaître trois langues : allemand, français et luxembourgeois.

Rares sont les sites institutionnels des banques qui ne proposent pas des informations en anglais. En France, le portail de l'Administration française (Service-Public.fr) donne accès à une partie de son contenu aussi en allemand, en anglais et en espagnol.

L'entreprise monolingue est une entreprise multilingue qui s'ignore. L'entreprise et les organisations monolingues sont en effet souvent confrontées à la nécessité de devenir multilingues. On traduit pour des publics divers : citoyens nationaux, résidents étrangers, touristes, étudiants, etc. Une question néanmoins se pose : comment considérer les expressions à traduire qui sont souvent propres à l'entreprise ou à l'organisation et qu'on a du mal à faire appartenir à la langue ? On tentera de répondre à cette question en deux étapes : d'abord par rapport à la lexicologie, ensuite par rapport à la terminologie.

2 Les mots en entreprise : un lexique à part ?

Les mots en usage dans l'entreprise constituent un lexique particulier. « Particulier », d'abord, parce que, en rapport avec la langue générale, il répond aux besoins des entreprises et non à celui de tous les locuteurs potentiels de la langue. Ensuite, parce qu'il s'actualise dans le discours des usagers, des clients et des employés par rapport à l'entreprise particulière et unique qui le crée et non par rapport à une autre. Ce lexique serait susceptible de passer dans la langue générale lorsque sa référence à l'entreprise n'aurait plus lieu d'être (ex. : *kleenex*, *frigo*). Voyons quelques exemples dans les paragraphes suivants.

Les expressions « horaires bleus » et « horaires verts » sont compréhensibles et traduisibles en soi, ce sont des < types d'horaire >, mais elles ne signifient quelque chose que si elles sont rattachées à l'utilisation que la Régie des transports de Marseille (RTM) en fait.

Il en va de même pour le « conseiller régional » ou le « conseiller en exploitation » chez McDonald's où ils désignent des postes de travail spécifiques.

L'« équipier de ventes » renvoie, chez Carrefour, au < chargé du réapprovisionnement des rayons >. Le syntagme semble maintenant d'ailleurs s'être diffusé rapidement grâce aux sites d'intérim, si l'on croit aux résultats proposés par *Google*.

Le « cadologue », où l'on reconnaît les formants du mot-valise : *cadeau* et *-logue* de *catalogue*, est le < catalogue des cadeaux > que l'on peut obtenir dans le supermarché Champion grâce aux points accumulés dans le < programme de fidélité > de l'enseigne et utilisables seulement dans les supermarchés Champion qui ne sont pas encore au 1^{er} septembre 2009 devenus Carrefour Market, le groupe Carrefour ayant acquis la chaîne Champion.

Le « bulletin de retard » est, à la Société nationale des chemins de fer français (SNCF), polysémique : c'est autant le petit < document en papier que l'on remet à un voyageur en gare ou édité via internet > que le < service d'information sur les retards qui ont lieu du lundi au vendredi de 6h à 9h du matin >.

Le « nanofoude » (nom masculin d'après les documents écrits chez idTGV en France) constitue une gamme des « petites bouchées gourmandes » – telles qu'idTGV les présente – ou des < grignotises > selon l'usage de plus en plus répandu dans la langue courante et en vente chez les trains idTGV.

Le « SéniorCité » est un < programme dédié aux plus de 60 ans > de la Régie autonome des transports parisiens (RATP).

Le « forfait classique 4 heures » est le < forfait de référence payé pour un service de communication par téléphonie mobile d'une durée de quatre heures > chez Bouygues Télécom en France.

Ces expressions se fondent dans le français courant tout en étant forgées par l'organisation ou l'entreprise qui les a fait connaître. Il est impossible de travailler sans ce lexique. Il est au service de l'identité (« C'est un produit à nous »), de la connaissance (il faut savoir ce qu'on vend ou produit) et de la culture de l'institution (« Chez nous, on fait comme ça », ce qui se fait ici peut ne pas se faire là-bas). De plus, les symboles® ou™ qui signalent la protection juridique de ce lexique n'est pas systématique.

Ce lexique n'est pas interchangeable avec celui d'une entreprise comparable. S'agissant des établissements de transports, les < horaires verts > de la RTM ne correspondent à aucune notion à la RATP. Les < conseillers régionaux > de McDonald's n'ont pas d'équivalent chez Quick où l'on a affaire à des « postes d'équipier », des « postes de manager » ou des « postes de directeur » pour un concept qui reste à identifier si on compare les deux enseignes. Le programme < SéniorCité > de la RATP correspondrait – le conditionnel est important – à la « carte OPUS » de la Société de transport de Montréal au Québec, où l'âge de référence pour en bénéficier est de 65 ans. Du côté extérieur de l'entreprise et si l'on veut profiter de ses bénéfices, il faut savoir ce qu'est le programme < SéniorCité >. À l'intérieur de la RATP, c'est tout un réseau d'expressions qui viennent soutenir ce programme, réseau auquel on accède que si l'on travaille à la RATP et, qui plus est, pour ce programme précis, parmi d'autres. On entrevoit alors que, pour chaque entreprise, un lexique particulier se constitue en un réseau qui n'est pas visible de l'extérieur mais qu'il est nécessaire d'acquérir pour intégrer l'entreprise et pour y travailler efficacement. Ces expressions deviennent des véritables unités d'information dont le contenu (trop particulier et sectoriel) n'intéresse pas la langue générale.

Néanmoins, ce sont bien des mots qui sont à l'œuvre s'écartant de l'utilisation habituelle prévue par les dictionnaires de langue où d'ailleurs ils n'auraient pas leur place. Le « cadologue » ou le « nanofoude » sont correctement formés phonologiquement, morphologiquement, s'insèrent normalement dans les phrases où ils sont nécessaires, mais leur accorder un statut de néologisme ne nous paraît pas approprié dans la mesure où ils peuvent disparaître si l'entreprise n'en a plus besoin². À partir du moment où ils sont créés par ceux qui entreprennent une activité (publique ou privée), nous proposons de les appeler « termes maison » pour signaler ponctuellement leur existence du point de vue lexicologique. L'histoire dira si – pour certains d'entre eux – ils deviendront des futurs « velcros ». En tout cas, leur datation trace l'histoire et l'évolution de l'entreprise, ce qui est un élément considérable du point de vue étymologique.

S'agit-il du lexique de la langue générale? Une réponse affirmative serait extrêmement hasardeuse. D'ailleurs, l'enseigner en FLE serait peu approprié. Du point de vue de l'apprentissage de la langue, nous avons déjà remarqué que nous serions

² C'est le cas quand une gamme de produits change de nom. À la Société de transports de Montréal, la « carte Opus » a remplacé la « carte d'Or ». Être attentif à l'oral est très important, c'est alors qu'on entend un employé d'un magasin de bricolage dire: « Ça n'existe plus, mais il faut chercher l'équivalent dans la nouvelle gamme. C'est parce qu'ils lui ont changé le nom! »

plus proches du français pour objectifs spécifiques (FOS) (de Vecchi, 2006 : 239). Et pourtant, ce lexique est présent au quotidien, dans les exemples des horaires, dans la langue de tous les jours, dans la parole des usagers et des prestataires de services de transport. L'évincer de la description de la langue en situation de travail est impossible, notamment par les intéressés qui ont besoin de travailler et qui ne peuvent s'en passer. L'entreprise aurait-elle ses mots à elle ? Si la réponse est positive, le lexicologue a besoin d'une classification adaptée, d'où notre proposition d'appellation : « terme maison » pour les mots issus de la créativité lexicale des acteurs des entreprises et pour les besoins de celles-ci. Quant au lexicographe, les termes maison seraient à considérer selon les besoins du répertoire à rédiger.

Ce lexique n'a donc pas sa place dans un dictionnaire de langue usuel ; tout en étant dans la langue, il n'y est pas. Mais à partir du moment où il se constitue en réseau, la terminologie offre d'autres perspectives, notamment celle d'exploiter et de traduire ce qui est conceptualisé et rendu par des termes ; dans ce cas, la notion de « mot » n'est plus opérationnelle.

3 Des mots en entreprise aux termes en entreprise

La langue se spécialise pour énoncer une connaissance, et davantage pour énoncer la connaissance spécifique d'une entreprise, voire son savoir-faire. De ce fait, sa terminologie acquiert une valeur ajoutée et a des implications multiples. Irremplaçables et difficilement échangeables, ces expressions irriguent tout le tissu de l'entreprise au point qu'on peut les considérer comme faisant partie de ses actifs à la fois matériels (une base de données a une valeur marchande) et immatériels (ce qu'elles encapsulent a une dimension qui va au-delà de la seule forme linguistique et peut également être marchande) des entreprises. Le matériau linguistique auquel nous nous attachons est difficile à saisir par l'observateur. Trop vite identifié comme un jargon, ce matériau est négligé ou souvent peu valorisé par les locuteurs. Nous l'avons appelé « parler d'entreprise » (de Vecchi, 1999 : 316) et dans son noyau se trouvent des termes dont les expressions que nous avons vues jusqu'ici sont des exemples.

Comme pour les mots vis-à-vis de la langue, les termes sont dans la langue spécialisée, tout en n'y étant pas, nous l'avons dit. La terminologie d'un domaine n'a pas vocation à inclure des parlars d'entreprise, cas particulier de langue spécialisée. Et pourtant, des termes comme *bulletin de retard* ou *SéniorCité* font partie de la langue spécialisée dans les transports. Or un découpage adapté de la notion de domaine permet de résoudre l'apparent paradoxe. Si nous considérons que l'entreprise exploite une activité dans un domaine de connaissance, nous pouvons alors comprendre que les besoins d'exploitation amènent les acteurs des entreprises à créer les termes qui, en reflétant leurs conceptualisations, les distinguent de leurs concurrents tout en remplissant une fonction utile (*cf. utilitas* de Lucrèce). De cette manière, le domaine de connaissance inclut des domaines d'activité (limite de la normalisation externe et début de la normalisation interne de l'entreprise) et ces

derniers, des domaines d'exploitation (de Vecchi, 2005 : 73). C'est bien pour cette raison que deux entreprises (domaine d'exploitation) – chacune avec son propre parler d'entreprise – qui fusionnent dans un même secteur (domaine d'activité) ne partagent pas les mêmes expressions, leurs parlers laissant apparaître, entre autres, des problèmes de communication. D'ailleurs, leurs conceptualisations ne sont pas forcément les mêmes. Autrement dit, leurs manières d'appréhender un marché différant, en conséquence, leurs réseaux notionnels à l'œuvre ne se recouvrent pas et leurs termes n'auront pas les mêmes implications. L'approche pragmateterminologique a l'ambition de rendre compte de cet ensemble d'implications, nécessaires autant au traducteur qu'à l'entreprise elle-même.

Le traducteur et l'entreprise ont besoin de références. Deux raisons viennent appuyer cet argument. La première est de faire en sorte que le client destinataire de la traduction puisse comprendre le discours de l'entreprise et avoir des points de repère³ d'abord en situation monolingue, ensuite en situation plurilingue. Pour cette raison, les glossaires, vocabulaires et lexiques fleurissent pour donner les définitions des termes proposés par l'entreprise, même si elles font l'objet d'une normalisation par ailleurs. Il s'agit de s'assurer que le client comprend ces termes dans le sens où l'entreprise les emploie. La deuxième raison est que ce matériau linguistique visible de l'extérieur émerge de l'intérieur de l'entreprise. L'analyse du fonctionnement interne de l'entreprise laisse apparaître (et en faisant une synthèse très serrée) que quatre fonctions essentielles des entreprises sont concernées. Les ressources humaines (*in fine* ce sont bien des humains qui communiquent), le marketing (conçu ici comme la réponse à un marché donné), la finance (il faut bien des capitaux pour financer ces activités) et la production (l'entreprise fabrique un produit ou propose un service). Dans ces fonctions, ce sont des acteurs qui conceptualisent, le résultat étant justement l'apparition de termes destinés à circuler souvent et dont la signification doit être partagée par tous. La perspective pragmateterminologique se démarque radicalement de celle de la lexicologie. Le terme existe en effet comme le résultat d'un assemblage de conditions données par les acteurs. C'est pour des raisons onomasiologiques qu'il est produit.

L'approche pragmateterminologique est légitime dans le traitement des termes d'un parler d'entreprise pour cinq raisons. La première est d'ordre linguistique : les termes expriment ce que les acteurs conçoivent. La deuxième est d'ordre cognitif : ils contribuent à la connaissance de l'entreprise. La troisième est d'ordre social : ils circulent dans des réseaux à l'intérieur et à l'extérieur de l'entreprise (ce que l'utilisateur ou le client voit). La quatrième est liée au temps : les termes de l'entreprise naissent, évoluent et meurent. À titre d'exemple, le « forfait classique 4 heures » de Bouygues Télécom s'appelait *forfait 240* avant le passage à l'euro en 2002⁴.

3 Par exemple l'organisme de crédit Cetelem en France qui propose un « Lexique crédit à la consommation ». L'entreprise Shell est bien plus précise encore dans la manière de faire interpréter les expressions contenues dans son rapport annuel de l'année 2007.

4 Il correspondait à 240 francs français et à 240 minutes de communication mensuelle. La motivation du terme alors évidente avait disparu lors du passage à l'euro (240 FF = 36,59 €). Le terme « Forfait 36,59 € » aurait eu peu de chances de s'implanter auprès des clients.

Quant à la cinquième raison que nous développerons plus largement, elle est d'ordre pragmatique. Elle tient notamment au rôle des verbes dans les collocations. D'une part, ces verbes renseignent sur la nature des actions, états et processus qui se développent dans l'entreprise, d'autre part sur la manière de les exécuter. La culture de l'entreprise se laisse entrevoir dans ses faits et gestes. Ces implications culturelles apparaissent de façon plus marquée dans le cas des fusions acquisitions: « chez nous, ça se fait comme ça ». Par rapport aux actions concrètes à effectuer, on peut citer, par exemple, le fait qu'un ticket de transport se paye, s'achète, se poinçonne, se vend, se délivre, se présente en cas de contrôle, se valide, se compose (pratique de la SNCF). Ce sont des collocations souvent prédictibles, voire évidentes, cependant il existe parfois des collocations inattendues dans un parler d'entreprise. Il importe de les considérer parce qu'elles ont été forgées par les acteurs d'une entreprise particulière à partir de sa manière de concevoir un produit ou un service. Par exemple: à Marseille, le « ticket SOLO » de la RTM bénéficie de la « correspondance Libertés » et donne accès à certains parkings de la ville de Marseille. Cette collocation n'est pas transposable pour toutes les compagnies de transport. Par ailleurs, le parler d'entreprise se lit, mais il s'entend aussi. L'étude de l'oral dans les organisations offre des perspectives pour l'étude des collocations, lesquelles informent sur le rôle du terme dans la communauté qui l'utilise et, au-delà, dans l'univers notionnel de l'entreprise. La gestion de la connaissance en entreprise ne peut faire l'économie de l'observation de l'oral, fournisseur important d'informations.

Derrière les termes des parlers d'entreprise se cache donc une dimension bien plus vaste que la dimension lexicologique: celle du réseau et des implications qu'ils créent dans le temps et dans l'action. Est-ce que Carrefour maintiendra le « cadologue » lorsque l'enseigne Champion aura disparu? Rien n'est moins sûr car chez Carrefour la < carte de fidélité > correspondant au < programme de fidélité > s'appelle « Ma Carte de Fidélité » et non « carte Champion » et les conditions d'utilisation sont différentes: des cadeaux chez Champion, des < avantages > chez Carrefour qui se concrétisent en « Avantages Fidélité » tels que des chèques à faire valoir dans les magasins participants et aux conditions d'utilisation spécifiques. Plusieurs entreprises utilisent l'expression « cadologue », mais les contenus notionnels ne sont pas les mêmes chez Champion et chez Igo-Post, entreprise qui l'a d'ailleurs enregistré comme marque en le faisant suivre d'un ®. Christophe Roche écrit: « *In fine* il est important de souligner que les structures lexicales et conceptuelles [...] ne se superposent pas avec la conceptualisation du monde: dire n'est pas concevoir. » (2007: 9) En entreprise, lorsqu'on conçoit, on termine par dire pour ensuite faire – sinon il n'y a pas d'entreprise, mais ce n'est pas parce qu'on dit que l'on conçoit la même chose. La comparaison entre entreprises d'un même secteur montre que les réseaux notionnels ne sont pas équivalents et qu'en conséquence il n'y a pas de raison pour que, de leur côté, les termes qui les reflètent soient les mêmes.

Que faut-il traduire: le terme de l'entreprise qui comporte tous les aspects que nous venons d'étudier ou bien ce qui est nécessaire à l'entreprise de véhiculer comme message dans d'autres langues pour être sûr de capter et de fidéliser des clients et, selon les cas, des employés? La réponse vient de l'entreprise ou de l'organisation

elle-même, non parce qu'elle commande la traduction, mais parce qu'elle sait (en principe) ce qu'elle veut obtenir comme résultat concret et évaluable.

Conclusion

De manière consciente ou inconsciente, l'entreprise sait que son langage véhicule des informations, parfois confidentielles et stratégiques; autrement dit, la considération et la protection de ses données sont un enjeu majeur dont elle doit être consciente. Or l'entreprise, à ce jour, est un terrain encore trop peu fréquenté des linguistes. Lorsque les linguistes y sont sollicités, c'est à huis clos: on ne divulgue pas des secrets de fabrication (dans le sens propre comme figuré). À l'instar du consultant, le terminologue ou l'ontologue travaillent avec des clauses de confidentialité. Les bases terminologiques ou les ontologies internes ne sont que rarement publiées. Hors cadre légal, la normalisation linguistique n'est pas une priorité pour les entreprises. Si une normalisation linguistique interne existe, c'est certainement pour des besoins de communication. En allant au cœur de l'entreprise, et dans des situations unilingues ou multilingues, on voit que les mots cèdent le pas aux termes et que, pour les observer, il faut tenir compte d'aspects qui vont au-delà du traitement terminologique traditionnel.

Pour une entreprise comme McDonald's ce qu'il importe de traduire n'est pas le terme « conseiller en exploitation », mais de rendre accessible dans une autre langue la notion: < conseiller en exploitation >, d'où l'importance fondamentale de ce que les acteurs conçoivent, puis expriment. L'économie de ce mécanisme ne peut être faite, parce que ce qui doit être traduit comme nous l'évoquions plus haut est le positionnement de l'entreprise, sa place dans le monde économique ou dans la société, sa politique propre.

La traduction n'est pas une anecdote, mais une raison essentielle pour que les entreprises voient la langue comme une ressource et non plus comme un simple moyen. Si ce n'est pas le cas, on verra encore des exemples malheureux comme celui de Motorola (et ce n'est pas un exemple isolé) qui avait lancé en France, en août 2008, le téléphone mobile « Mon Q » (*sic*) avant de faire les modifications nécessaires et le transformer en « Mon Q9h » (*sic*) tel que le montrait le site internet de Motorola en août 2009.

Pour conclure, nous dirions que le terminologue et le traducteur, passeurs de mots, tiennent des rôles dans le traitement des parlers d'entreprise et que ces rôles ne sont pas des moindres. C'est au terminologue d'assurer la cohérence conceptuelle (et communicationnelle) des termes des parlers d'entreprise. Quant au traducteur, c'est à lui d'assurer – pour ne pas dire de garantir – que l'entreprise garde sa cohérence en passant d'une langue à une autre. Nous pensons que dans la lutte économique acharnée actuelle, plus que jamais l'entreprise gagnerait à avoir des terminologues et des traducteurs maison, imprégnés de sa culture et participant dans la mesure du possible autant à ses ressources humaines qu'à sa communication

interne et externe. L'entreprise ne parle pas. Elle n'écrit pas non plus. Ce sont ceux qui, en travaillant, conçoivent, parlent et écrivent pour communiquer. C'est la raison pour laquelle la langue en entreprise et sa remise en question frontale ne sont pas anecdotiques.

Bibliographie

- Jardel (J.-P.) et Loridaon (C.), 2000 : *Les rites dans l'entreprise*, Paris, Editions d'Organisation.
- Lehnisch (J.-P.), 1991 : *La communication en entreprise*, 3^e éd., Paris, PUF, Coll. QSJ n° 2229.
- Lerat (P.), 1995 : *Les langues spécialisées*, Paris, PUF.
- Lucrèce, éd. 1971 : *De la nature*, T. II, Livres IV – VI, V, 1028-1029. trad. Alfred Ernout, Paris, Belles Lettres.
- Michaud (C.) et Thoenig (J.-C.), 2001 : *Stratégie et sociologie de l'entreprise*, Paris, Village mondial.
- Thévenet (M.), 1999 : *La culture d'entreprise*, 3^e éd., Paris PUF, Coll. QSJ n° 2756.
- Roche (C.), 2008 : « Le terme et le concept : fondements d'une ontoterminologie » dans *TOTH 2007*, C. Roche éd., Annecy, Institut Porphyre, p. 1-22, www.porphyre.org.
- (de) Vecchi (D.), 2007a : « De la culture d'entreprise et de la réalité linguistique de l'entreprise », dans Brémond (J.) et Massol (J.), éd., *Entreprise, cultures nationales & mondialisation, Actes du colloque international et interdisciplinaire, Université de Nantes, 6 et 7 décembre 2007*, Nantes, CIL – CRINI, p. 287-303.
- (de) Vecchi (D.), 2007b : « Terminologie et sciences de gestion. Le cas des entreprises : vers une pragmatoterminologie » dans *Terminologie : approches transdisciplinaires, Gatineau, UQO, Québec 2-4 mai 2007*, actes en ligne : www.uqo.ca/terminologie2007/documents/deVecchi.pdf.
- (de) Vecchi (D.), 2006 : « Place de la terminologie dans l'enseignement des langues spécialisées dans une école de management », dans Bertrand (O.) et Schaffner (I.), éd., *Apprendre une langue de spécialité : enjeux linguistiques et culturels, Actes du colloque international, École polytechnique, 14-15 septembre 2006*, Paris, Presses de l'École polytechnique, p. 237-249
- (de) Vecchi (D.), 2005 : « La terminologie dans la communication de l'entreprise, approche pragmatoterminologique », dans *Cahiers du CIEL*, Université Paris 7 EILA, mars 2005, p. 71-83.
- (de) Vecchi (D.), 2002 : *La terminologie en entreprise. Formes d'une singularité lexicale*, Thèse de doctorat, Paris, Université Paris 13.
- Véry (P.), 2002 : *Des fusions et des hommes*, Paris, Editions d'Organisation.

Sources sur internet

- Bouygues Télécom*, 2009 : www.laboutique.bouyguestelecom.fr/nos-offres.html.
- British Museum*, 2009 : www.britishmuseum.org/, consulté le 5 juillet 2009.
- Cityrama*, 2009 : www.pariscityrama.com/, consulté le 5 juillet 2009.
- Cetelem*, 2009 : « Lexique crédit à la consommation », www.cetelem.fr/credit/guide-credit/lexique.php, consulté le 5 juillet 2009.
- Coopérative Eroski*, 2009 : www.eroski.es/, consulté le 5 juillet 2009.
- École de management de Marseille*, 2009 : www.euromed-management.com/, consulté le 3 août 2009.

École supérieure de commerce de Paris Europe (ESCP Europe), 2009 : www.escpeurope.eu/, consulté le 3 août 2009.

L'Express, 2009 : Carrières – Les clefs de la promotion interne: www.lexpress.fr/informations/carriere-les-clefs-de-la-promotion-interne_619842.html, consulté le 20 juillet 2009.

Groupe Carrefour, 2009 : www.carrefourmyjob.com/groupe/enseigne_carrefour.htm, consulté le 4 septembre 2009.

idTGV, 2009 : www.idtgv.com, consulté le 5 juillet 2009.

Magasins Champion, 2009 : www.champion.fr, consulté le 5 juillet 2009.

Magasins Darty, 2009 : www.darty.lu/, consulté le 3 août 2009.

Motorola, 2009 : direct.motorola.com/CFN/motoq9/q-home.asp?productid=30722, consulté le 3 août 2009.

Office du tourisme de Paris, 2009 : www.parisinfo.com/, consulté le 5 juillet 2009.

(Le) Portail de l'Administration française, 2009 : www.service-public.fr/, consulté le 3 août 2009.

RATP, 2009 : www.ratp.fr, consulté le 5 juillet 2009.

Régie des transports de Marseille, 2009 : www.rtm.fr, consulté le 5 juillet 2009.

Shell, 2008 : « Rapport annuel 2007 », www.annualreportandform2of.shell.com/2007/servicepages/about_disclaimer.php, consulté le 4 août 2009.

SNCF, 2009, Mon bulletin de retard: www.ter-sncf.com/Regions/centre/Fr/Se_deplacer_en_TER/Apres_mon_voyage/Bulletins_de_retard/Default.aspx, mise à jour le 6 novembre 2009.

Société de transports de Montréal: www.stm.info/tarifcation/65ans.htm, consulté le 5 juillet 2009.

Supermarchés Migros , 2009 : www.migros.ch/, consulté le 5 juillet 2009.

Velcro: www.velcro.fr, consulté le 5 juillet 2009.

Jeanne Dancette,
Département de linguistique et de traduction,
Université de Montréal,
Montréal,
Canada,
Jeanne.dancette@umontreal.ca.

Un dictionnaire encyclopédique plurilingue sur thésaurus¹

Résumé

Cet article discute de l'intérêt et de la faisabilité d'un dictionnaire combinant les traits d'une encyclopédie et d'une ontologie. Le cadre de cette discussion est la gestion des connaissances, leur partage et leur exploitation en situation de traduction/rédaction, ou plus généralement d'acquisition de connaissances. Des articles de dictionnaires existants sont analysés du point de vue de leur valeur comme outil de connaissance. Le *Dictionnaire analytique de la mondialisation du travail* (DAMT) est présenté afin de démontrer que les modèles encyclopédique et théaural peuvent être combinés pour répondre aux besoins de documentation et de recherche lexicale des utilisateurs. L'article insiste sur l'exploitation des relations sémantiques et conceptuelles, et replace au centre de la discussion l'utilisateur qui devient le gestionnaire de ses connaissances grâce aux différents modes d'accès à l'information.

Mots-clés: encyclopédie; thésaurus; terminologie plurilingue; base de connaissances terminologique; relations sémantiques et associatives; gouvernance.

Introduction

Le projet que nous présentons porte sur la conception d'un outil dictionnaire multilingue qui réponde aux besoins de documentation et de recherche lexicale des utilisateurs, en premier lieu, des traducteurs spécialisés. Le concept dictionnaire que nous favorisons est à cheval entre celui de l'encyclopédie spécialisée et celui de l'ontologie. Il se présente comme une base multilingue de connaissances terminologiques.

Nous en exposerons les traits principaux en nous appuyant sur des exemples d'ouvrages illustrant la méthode et dégagerons l'originalité de notre modèle dictionnaire qui met en valeur un système développé de relations sémantiques et

¹ Nous souhaitons souligner la contribution de notre assistante de recherche, Rosa Castrillon, à l'élaboration des fiches terminologiques et à la mise en page de cet article. Nous remercions également le Centre de recherche en sciences humaines (CRSH) du Canada pour le financement de la recherche.

associatives. Pour finir, nous replacerons au centre de notre discussion l'utilisateur du dictionnaire qui devient lui-même le gestionnaire de ses connaissances.

1 Cadre théorique

L'ouvrage que nous présentons aujourd'hui (Dancette 2009a, version de travail) a été conçu pour répondre aux besoins de documentation (informer sur les concepts) et de recherche lexicale (préciser le sens et l'emploi des termes et des collocations spécialisées, dans plusieurs langues). Il a pour public cible les traducteurs, étudiants, rédacteurs dans le domaine de spécialité visé, ici la mondialisation économique et ses effets sur le travail. Cette communauté est large et hétérogène, mais rassemble les personnes intéressées aux questions de relations industrielles, d'économie et de droit du travail, que ce soit pour des questions terminologiques ou de connaissances sur les notions.

Le cadre théorique général dans lequel s'inscrit cette recherche est celui de la gestion des connaissances. Les dimensions qui l'orientent sont notamment : 1. le partage des connaissances dans un univers multidisciplinaire où le rédacteur de dictionnaire et les experts dans le domaine travaillent en synergie ; 2. l'organisation des connaissances dans un environnement de langue naturelle où la logique des relations entre concepts répond à une méthode de recherche heuristique, de proche en proche ; et 3. leur exploitation et leur extraction en situation soit de traduction soit de rédaction, soit plus généralement d'acquisition de connaissances dans un domaine spécialisé. Entendue dans cette acception, la gestion des connaissances est une application pratique des sciences cognitives qui en sont le fondement théorique.

Notre modèle dictionnaire se conçoit comme une base de connaissances terminologique (BCT) (Otman 1997). Par définition, une BCT contient des informations terminologiques et une description explicite et structurée des concepts sous forme de « réseaux sémantico-terminologiques ». Le dictionnaire présenté constitue un vaste réseau de termes unis par des relations sémantiques (RS) et des relations associatives (RA) reflétant l'univers conceptuel du domaine. L'étiquetage du lien entre deux unités terminologiques en est une particularité importante et originale : il permet une certaine formalisation de l'information qui en rend plus faciles tant la compréhension que la récupération.

Le réseau de termes où chaque fiche renvoie à une autre fiche dessine un circuit de connaissances et répond à une forme souple, modulable et évolutive du « savoir » :

« Le savoir affecte forcément une forme circulaire : c'est en effet la seule manière de se représenter un ensemble de données diverses tel que chacune renvoie à toutes les autres et ait perspective sur toutes les autres. [...] Ce savoir n'est pas simplement cumulatif [...] mais circulaire parce qu'il y a une circulation du savoir d'un point quelconque à tout autre point possible. Sans doute cette circulation se fait-elle le long de certains axes perspectifs qui seront par habitude plus fréquentés que

d'autres à l'intérieur du tout, mais dont la commodité ne tient jamais finalement qu'à un état momentané du savoir, à un équilibre météorologique métastable. » (Varet 1956 : 69)

Gardant à l'esprit le principe de circularité du savoir, la question posée par notre modèle dictionnaire est celle de la praticité de l'outil pour l'utilisateur qui cherche à orienter sa recherche et à devenir son propre gestionnaire de l'information. Nous voulons démontrer la valeur aux fins de la recherche d'information d'une organisation de connaissances sous forme de descriptions en langue naturelle, d'une part, et en langage semi-formel, d'autre part, via l'explicitation des relations entre termes.

2 Concept dictionnaire

Le concept dictionnaire que nous avons privilégié combine les caractéristiques d'une encyclopédie et d'une ontologie. Une analyse comparative de ces deux types d'ouvrages permet de mettre en évidence ce qu'ils ont en commun et ce qui les oppose.

2.1 L'encyclopédie et l'ontologie, deux types d'ouvrages compatibles ?

Il s'agit de types classiques d'ouvrages visant l'exploitation et l'organisation des connaissances. Ils ont en commun la description d'un champ du savoir humain et tous les deux s'inscrivent dans une logique d'accumulation du savoir. Mais, ils s'opposent par d'autres aspects.

- L'objectif de l'encyclopédie est la transmission d'un ensemble de connaissances ; celui de l'ontologie est le repérage de l'information afin de faciliter, dans la plupart des cas, une prise de décision ;
- Ils divergent par leur mode rédactionnel. L'encyclopédie (de celle de Diderot et d'Alembert (1751-1765) à *Nupedia*, ou à *Wikipédia*, en passant par *Universalis*, *Encarta*, le *Quid*, *Britannica*, etc.) reproduit le discours savant, en langue naturelle, sous forme déclarative. L'ontologie ne renvoie pas à un discours ; elle exploite une écriture en réseau (les entités sont toutes reliées), sous une forme procédurale ;
- Dans le cas de l'encyclopédie, les articles sont validés par des experts et les auteurs recrutés pour les rédiger font autorité. Ils visent une certaine universalité dans le choix de l'information retenue. Par opposition, les entrées de l'ontologie sont anonymes (généralement) et non validées. Ses concepteurs ne se donnent pas l'objectif de créer des réseaux de termes universels, valables dans tous les contextes d'utilisation ;
- Enfin, en ce qui concerne le mode d'accès à l'information, la navigation par hyperliens est un élément essentiel de l'ontologie, mais facultatif dans le cas de l'encyclopédie.

² Certaines des distinctions que nous faisons se retrouvent dans Ertzcheid (2005) qui oppose Wikipédia et KNOL.

Nous illustrons cette opposition à partir d'articles réels tirés d'ouvrages existants et aussi du DAMT (Dictionnaire analytique de la mondialisation du travail/Analytical Dictionary of Globalization and Work/Diccionario analítico de la globalización del trabajo — Dancette, version 2009).

2.2 L'encyclopédie

On trouvera à l'annexe 1 l'article GOUVERNANCE, extrait de l'*Encyclopédie Wikipédia*.

Dans cet article, nous observons les caractéristiques que nous relevions ci-dessus : le recours au mode discursif du texte spécialisé, approprié pour faire passer une information réputée juste et contrôlée. *Wikipédia* explique ainsi son mode de validation :

« Le dernier niveau de contrôle, cette fois, collectif, est formé par les programmes d'amélioration systématique de tous les articles rattachés à un thème donné (organisé autour d'un « portail »). Dans ce cadre, les articles sont systématiquement relus, complétés et corrigés, par des équipes de volontaires passionnées par ce thème et le plus souvent d'un bon niveau de spécialisation. Les articles qui bénéficient de ces relectures sont initialement corrigés, et continuent généralement à être suivis par l'équipe du « portail ». De plus, sur un article complet et cohérent, il est moins tentant d'ajouter des erreurs, qui se voient d'autant plus facilement. » (Wikipédia 2009)

La circulation d'un article à l'autre selon les besoins particuliers de l'utilisateur est rendue aisée grâce à la création des hyperliens. Dans l'article pris en exemple, l'utilisateur est renvoyé aux spécifiques GOUVERNANCE D'ENTREPRISE, GOUVERNANCE POLITIQUE, GOUVERNANCE RÉGIONALE, GOUVERNANCE MONDIALE, GOUVERNANCE LOCALE.

L'efficacité de *Wikipédia* comme outil de transmission de la connaissance n'est plus à démontrer malgré les critiques que suscitent certains articles qui font l'objet de canulars ou qui échappent au contrôle éditorial. Sa large diffusion indique aussi le besoin auquel il répond³.

2.3 L'ontologie : quelques exemples

Une ontologie (le thésaurus⁴ en étant un type particulier), par opposition à l'encyclopédie, est avant tout un réseau de termes (ou simples dénominations d'entités) qui regroupe un ensemble de concepts décrivant un domaine. Les concepts sont liés les uns aux autres par des relations taxonomiques (hiérarchisation des concepts) d'une part, et des relations dites sémantiques ou simplement concep-

³ *Wikipédia* fait partie en 2008 des 10 sites les plus visités du *World Wide Web*. En 2009, le nombre d'articles a dépassé les 11 millions, dont plus de 2 400 000 dans *Wikipédia* en anglais, plus de 700 000 dans *Wikipédia* en allemand et *Wikipédia* en français.

⁴ Le thésaurus est une liste de termes. S'appuyant sur un langage documentaire, il est souvent défini comme un outil d'indexation. Il prend souvent la forme d'un dictionnaire hiérarchisé. Avec les relations sémantiques et les relations dites associatives que les logiciels de thésaurus permettent d'exploiter, les thésaurus se rapprochent des ontologies dans la mesure où les rapports entre les notions peuvent faire l'objet de descriptions très structurées.

tuelles⁵, d'autre part. Le vocabulaire rendant compte du domaine est contrôlé et structuré, mais les termes ne sont pas toujours définis et les relations entre les termes sont peu explicites, comme nous le verrons ci-dessous.

Bien que les appellations diffèrent, les auteurs distinguent plusieurs types d'ontologie : « statique », « dynamique », « orientée vers la tâche » (ou « intentionnelle ») et « sociale ». Celle qui nous rejoint par ses objectifs et ses modalités est l'ontologie statique, ainsi définie :

« [L]es ontologies statiques s'intéressent aux choses existant dans le monde – dans un monde donné – pour déterminer quels peuvent être leurs attributs et leurs relations. Il s'agit d'un type basique, qui ne prétend être ni universel ni trivial. » (Jurisica *et al.* 1999 : 15, traduit et cité dans Ertzcheid 2002 : 355).

Ainsi, dans une ontologie statique on consigne les dénominations d'entités et celles de leurs attributs et des concepts reliés sans tenir compte de variations qui peuvent affecter ces entités, par opposition à l'ontologie dynamique, orientée vers l'analyse des changements d'état ou de variables qui permettront de fonder les décisions (par exemple, un traitement dans le domaine médical).

Le format d'un dictionnaire qui vise le transfert de connaissances pour un large public ne correspond pas, de manière évidente, aux caractéristiques d'une ontologie dynamique⁶, ni intentionnelle⁷ ni sociale⁸, si l'on reprend la typologie et les définitions présentées par Jurisica *et al.* (1999). Dans ce qui suit, nous présentons quelques articles de dictionnaire ayant une structure ontologique (statique) et les analysons sous l'angle de leur valeur comme outil de connaissance.

2.3.1 L'article RÉGIME DE PENSION (Thésaurus du vieillissement et de la santé, 2003)

L'article est reproduit à l'annexe 2. Plusieurs observations sont à faire :

- Les entrées ne contiennent pas de définition au sens propre. Ce qui en ferait office, toutefois, est l'information donnée dans Note d'application, qui précise

5 Il semble inapproprié d'appeler « sémantiques » des relations qui relient entre eux des concepts ou des entités qui ne renvoient pas toujours à des termes. Les désignations mises en relation correspondent souvent à des concepts dont les rapprochements les uns avec les autres relèvent plus de la connaissance du monde que de la sémantique. Dans certains cas, il s'agit même de simples étiquettes d'entités qui ressemblent plus à des noms propres qu'à des termes. Dans l'exemple présenté dans le texte, le RÉER (régime d'épargne enregistré de retraite) est une réalité propre au Canada. Le fait que l'appellation RÉER a un référent unique ne la disqualifie-t-il pas en tant que terme ?

6 « A dynamic ontology, on the other hand, describes the changing aspects of the world in terms of states, state transitions and processes. » (Jurisica *et al.* 1999 : 7)

7 « Intentional ontologies encompass the world of motivations, intents, goals, beliefs, alternatives, choices, etc. Typical primitive concepts include issue, goal, supports, denies, subgoalOf, agent, etc. An intentional ontology allows alternate realities to be expressed and reasoned about. » (Jurisica *et al.* 1999 : 7)

8 « A social ontology covers social settings, organizational structures, or shifting networks of alliances and interdependencies (Galbraith, 1973; Mintzberg, 1979; Scott, 1987). Traditionally, social ontologies have been characterized in terms of concepts such as actor, position, role, authority, commitment, etc. Speech acts theory offers an ontology for modeling communication among actors (Medina & Mora, 1992). » Jurisica *et al.* 1999 : 7.

l'extension du concept générique RÉGIME DE PENSION et le statut de spécifique que présente RÉGIME DE RETRAITE.

- Quant aux relations entre les entités, on constate qu'elles sont peu explicites et renseignent peu sur les relations de sens. Elles sont de plusieurs types : l'équivalent linguistique (ici l'anglais) ; les synonymes ; les spécifiques et les « termes associés ». Dans la catégorie *Employé pour*, qui regroupe les synonymes, on retrouve, à côté de véritables synonymes comme FONDS DE PENSION ou PLAN DE PENSION, les termes PRESTATION et RENTE qui seraient à considérer plutôt comme génériques.

Parmi les Termes spécifiques se trouvent, comme il se doit, des types de régimes de pension (ex. : RÉGIME DE RETRAITE, PENSION D'INVALIDITÉ), mais aussi le terme ALLOCATION AU SURVIVANT qui pourrait être considéré comme une partie de RÉGIME DE PENSION, mais certainement pas comme un spécifique.

Les termes FINANCES PERSONNELLES et REÉR (régime enregistré d'épargne-retraite) ont été classés dans la catégorie Termes associés, sans que la relation conceptuelle ne soit précisée. Dans le cas de FINANCES PERSONNELLES, on pourrait penser qu'il s'agit d'une relation de partie-tout, puisque les finances personnelles incluent les régimes de pension. Dans le cas de REÉR, le rapport avec RÉGIME DE PENSION est lointain et mal défini : le REÉR est un programme d'incitation à l'épargne en prévision de la retraite.

On peut donc dire que les RS ne sont pas rigoureusement étiquetées et qu'elles n'offrent pas de véritables indices pour préciser le sens, l'extension ou le statut d'un concept par rapport à un autre. Quoi qu'il en soit, les entrées consignées dans le thésaurus permettent un repérage des concepts et une navigation d'un point à un autre à l'intérieur du domaine.

2.3.2 Thésaurus du BIT/Ilo Thesaurus

Nous examinons un autre exemple, le Thésaurus du BIT/Ilo Thesaurus (Bureau international du travail), dans le domaine des relations industrielles. Il s'agit d'un dictionnaire hiérarchisé multilingue (anglais, français, espagnol).

Comme dans l'article extrait du *Thésaurus du vieillissement et de la santé*, une définition est indiquée dans la rubrique SN (*Scope Note*), et comme dans l'autre ouvrage, les termes ne sont pas systématiquement définis (CORPORATE GOVERNANCE, par ex., n'est pas défini). Toutes les relations sémantiques se retrouvent dans la rubrique RT (*Related Terms*). Il s'agit de relations associatives floues dont la nature n'est pas précisée. Ainsi, on peut se demander quel est le lien entre GOVERNMENT et GOVERNANCE, ou entre LEGAL SYSTEM OU RESPONSIBILITY et GOVERNANCE.

Toutefois, malgré le côté peu développé des relations, cette compilation de 4 000 termes relatifs au monde du travail et de l'entreprise est très utilisée et fait autorité dans le domaine en tant que référence terminologique, pas seulement en vertu

du nombre de termes qu'elle valide grâce à l'autorité de l'OIT, mais aussi en vertu des définitions entérinées par cette organisation et qui, de ce fait, font consensus à l'échelle mondiale.

Ces deux exemples nous ont permis d'établir la limite des modèles théauraux traditionnels. Le flou dans l'établissement des relations, le faible nombre des relations et l'absence d'explicitation constituent une faiblesse reconnue (Tudhope *et al.* 2001) que nous comptons dépasser afin d'améliorer le modèle en tant qu'outil de connaissances.

2.3.3 Dictionnaire onomasiologique fondé sur une ontologie (Maroto 2007, Maroto et Alcina 2009)

Maroto (2007) a démontré avec force que la définition des relations conceptuelles et leur application dans un domaine très limité (les produits céramiques) sont faisables. Elle distingue cinq types de relations conceptuelles pour un total de 22 relations, par exemple, matériel/objet, étape/processus, lieu/action, instrument/procès, étape/procès, cause/effet, objet/usage, etc. Ce modèle très sophistiqué de relations conceptuelles démontre qu'il est possible de rendre compte d'un domaine d'activité humaine par toutes les caractéristiques et propriétés pertinentes pour la récupération d'une information précise. Il nous semble extrêmement intéressant pour sa valeur en tant qu'outil de connaissance.

Il devrait être possible de représenter tout domaine (y compris dans les sciences sociales, même si la tâche est plus complexe que dans un domaine technique) au moyen de relations conceptuelles étiquetées. Notre hypothèse est que le formalisme des relations sémantiques et conceptuelles permet un accès plus rapide à l'information que le texte descriptif. Il permet aussi de visualiser globalement un ensemble de concepts et de se le représenter logiquement en comprenant ce qui rapproche les concepts et ce qui les oppose.

Le modèle que nous avons développé et appliqué au domaine de l'économie mondiale n'enlève pas sa valeur au texte descriptif (encyclopédique). Bien au contraire, il vise à optimiser l'exploitation de l'information contenue dans les textes.

2.4 Un article du DAMT (Dancette, 2009a) illustrant le concept d'encyclopédie, de dictionnaire multilingue et d'ontologie

L'article GOUVERNANCE D'ENTREPRISE figure à l'annexe 3. Notons que les articles anglais et espagnol correspondant au même concept sont dans la même base, accessibles par hyperliens. (Il ne s'agit pas de traductions de l'article mais plutôt de textes parallèles qui constituent de ce fait des articles originaux, même si la structure reste fondamentalement la même.)

Un article se compose d'un texte (dans les rubriques Définition et Description) et de tableaux de relations sémantiques et conceptuelles qui font que chaque article est en réseau avec un ensemble d'autres articles. Nous allons surtout discuter de

cet aspect-là du dictionnaire, étant donné que la dimension ontologique fait l'objet principal de notre discussion.

Dans l'article donné en exemple, on trouve parmi les relations sémantiques : les relations d'équivalences interlinguistiques, les relations synonymiques et les variantes, ainsi que les quasi-synonymes, les génériques, les spécifiques, les antonymes (appelés contrastes)⁹. (On trouvera aussi dans certains cas des relations méronymiques (partie-tout) et des relations dérivationnelles, telles que GOUVERNER par rapport à GOUVERNEMENT.) Ce sont là des relations qu'il est habituel de trouver dans un dictionnaire multilingue.

Parmi les relations associatives ou conceptuelles, sont consignées les relations acteur (ACT) et législation (LEG), et des relations non étiquetées que nous appelons Concepts proches, pour désigner les relations qu'il est impossible de qualifier de manière formelle, comme RESPONSABILITÉ OU ÉTHIQUE DES AFFAIRES par rapport à GOUVERNANCE D'ENTREPRISE. Le lien entre les deux concepts est trop éloigné et trop instable pour être qualifié. (La liste de toutes les relations figure à l'annexe 6.)

Nous reproduisons ci-dessous le tableau des relations de GOUVERNANCE D'ENTREPRISE :

Relations sémantiques

| |
|--|
| Équivalents : <i>Corporate governance</i> <i>Gobernabilidad de la empresa</i> |
| Synonymes et variantes : <i>Gestion concertée de l'entreprise</i> , <i>Gouvernement de l'entreprise</i> , <i>Gouvernement d'entreprise</i> |
| Quasi-synonyme : <i>Régie d'entreprise</i> |
| Générique : <i>Gouvernance</i> |
| Contrastes : <i>Gouvernance syndicale</i> , <i>Gouvernance politique</i> |

Relations conceptuelles ou associatives

| |
|---|
| Acteur : <i>Actionnaire</i> , <i>Conseil d'administration</i> |
| Législation : <i>Charte de gouvernance d'entreprise</i> , <i>Code de bonnes pratiques de gouvernance d'entreprise</i> , <i>Code de gouvernance d'entreprise</i> , <i>Droit flexible</i> |
| Concepts proches : <i>Éthique des affaires</i> , <i>Responsabilité 1</i> |

Même si les termes *actionnaire*, *code de gouvernance d'entreprise* présentés dans les relations conceptuelles ne dénotent pas une caractéristique essentielle de la gouvernance d'entreprise, il y sont associés en tant qu'acteur ou document juridique

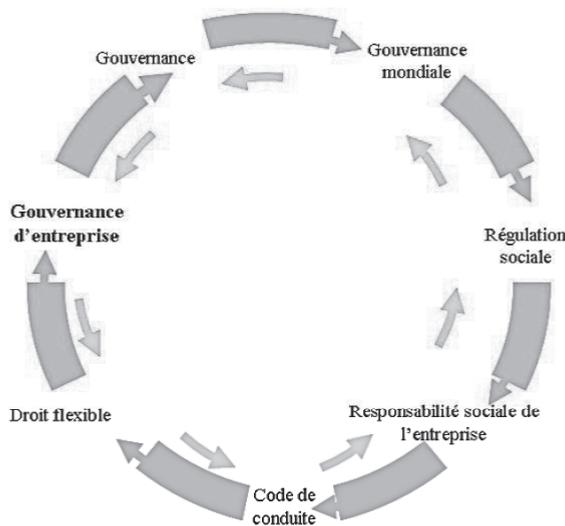
⁹ L'étiquette *Contraste* nous paraît plus précise que celle de *Antonyme* ou de *Contraire* pour rendre compte des nombreux cas où une entité s'oppose à une autre par une ou quelques-unes seulement de ses caractéristiques.

la concernant. Ces termes indiquent une relation conceptuelle, non hiérarchique et non syntagmatique (dérivationale), qui permet une extension de la compréhension du concept GOUVERNANCE D'ENTREPRISE. Ces termes reliés reçoivent le statut de fiche du simple fait qu'ils sont en relation avec *gouvernance d'entreprise*.

2.5 Circuit de navigation

Avant de poursuivre la discussion sur les relations, nous faisons une parenthèse sur leur intérêt sur le plan cognitif. Les liens implantés à l'intérieur d'une fiche renvoient à d'autres fiches. Ils illustrent l'idée de circularité d'un point du savoir à un autre (disons plus modestement, d'un point du dictionnaire à un autre), selon un axe perspectif, individuel et inscrit dans le temps.

Le graphe qui suit reproduit un circuit possible de navigation entre l'article GOUVERNANCE D'ENTREPRISE et les autres. De proche en proche, l'utilisateur peut arriver, en fonction de ses besoins, à l'article RÉGULATION SOCIALE OU DROIT FLEXIBLE.



Grappe 1. Circuit de navigation entre les concepts

3 Présentation de la base de données du DAMT

On trouvera à l'annexe 5 la reproduction de la page d'accueil du DAMT.

3.1 Modèle des relations sémantiques et conceptuelles dans le DAMT

À l'exception des relations indiquées comme Concepts proches, toutes les relations sémantiques sont définies et étiquetées. Cette caractéristique relève plus d'un héritage de la sémantique actantielle (Dancette 2007 et 2009 b), et notamment des fonctions lexicales de la Lexicologie explicative et combinatoire, que des modèles d'ontologies statiques qui se contentent généralement de quelques

relations conceptuelles, comme il a été démontré plus haut. À la différence des fonctions lexicales qui sont unidirectionnelles (à quelques exceptions près, telles que la fonction *CONVERSE*), dans un modèle ontologique ou thésaural, une relation exprime un rapport de réciprocité. Cela entraîne une contrainte forte : à toute paire de termes correspond une étiquette et une seule, quel que soit le contexte discursif dans lequel apparaît cette paire de termes.

Le logiciel de création de thésaurus *MultiTes* que nous utilisons pour produire le DAMT facilite la gestion de la nomenclature du dictionnaire. Il permet l'alignement des termes, la vérification de leur cohérence (uniformité de la terminologie) et la systématisation des étiquettes de relations en réponse à la contrainte ci-dessus énoncée. Mais nous avons dû enrichir les possibilités de *MultiTes* en créant un grand nombre de relations conceptuelles *sui generis*, afin de répondre à notre objectif premier : créer une base de connaissances exploitant l'information encyclopédique contenue dans les articles, comme démontré au point 2.4.

3.2 Nombre et nature des relations sémantiques et conceptuelles

Notre modèle compte 7 classes de relations sémantiques (équivalences inter- et intralinguistiques, relations hiérarchiques, méronymiques et syntagmatiques) et 7 classes de relations associatives. On trouvera en annexe 6 la liste de toutes les relations.

À ce stade, la base de données terminologiques comprend plus de 4 000 termes (dans les 3 langues confondues) dont près de 3 000 renvoient à des concepts distincts (les 1 000 autres étant des variantes ou des synonymes). Chaque terme fait l'objet d'une fiche avec définition et tableau de relations. La base dénombre près de 1 000 relations d'équivalence, environ 1 330 relations hiérarchiques (générique, spécifique, totalité, partie, singulatif et multiple), et près de 1 200 relations associatives de toutes sortes, dont une cinquantaine de relations syntagmatiques. Par ce nombre élevé de relations, le DAMT se distingue d'autres ouvrages dictionnaires plus classiques.

3.3 Modes de consultation de la base

L'utilisateur dispose de divers modes de navigation dans le DAMT :

- Du terme à la fiche. Les termes sont organisés et présentés par ordre alphabétique. On peut accéder à une fiche en parcourant l'index ou un sous-index (organisé par langue, ou par domaine). La consultation d'une fiche peut donc se faire en partant de tout terme de la nomenclature ;
- D'une fiche à une autre fiche. Des liens cliquables permettent de naviguer entre les termes reliés (à l'intérieur des tableaux de mots reliés) définissant le champ sémantique. Tous les mots reliés conduisent à une fiche ;
- D'une référence bibliographique à une page Web ;

- D'une référence d'organisme (adresse URL) à un site officiel (ONU, CISL, Union européenne). Cette fonctionnalité n'était pas prévue par *MultiTes*; il a fallu introduire manuellement des balises permettant le renvoi à une adresse externe (lien URL).

Conclusions

Le dictionnaire encyclopédique théssaural exploite une approche qui permet de mettre en évidence des systèmes notionnels complexes. Conçu comme une base de connaissances, il cherche à exploiter les avantages combinés d'un dictionnaire encyclopédique, d'un ouvrage terminologique multilingue et d'un thésaurus. Dans cette approche, la gestion des relations sémantiques est la clé de voute de la base (validité du réseau de connaissances).

Du point de vue de l'utilisateur, l'accès à l'information est subordonné aux modes de classement et d'organisation des données; ce qui permet une recherche personnelle orientée et non pas aléatoire, à la différence d'un moteur de recherche tel que *Google* ou *Yahoo*. Dans un contexte d'apprentissage ou de recherche heuristique, c'est un avantage appréciable. « Passeur de mots », le dictionnaire devient aussi passeur de savoirs et, ce qui est plus important encore, l'utilisateur devient le gestionnaire de l'information: il cherche ce qui est pertinent pour ses besoins.

Bibliographie

- Dancette (J.), 2009a : *Dictionnaire analytique de la mondialisation du travail*/*Analytical Dictionary of Globalization and Work*/*Diccionario analítico de la globalización del trabajo*, site intranet du Centre de recherche interuniversitaire sur la mondialisation et le travail (CRIMT), crimt.org.
- Dancette (J.), 2009b : « Des relations sémantiques comme balises de la connaissance; leur intégration dans des dictionnaires spécialisés multilingues », dans Quirion, (J.), Rousseau (L.-J.) et Depecker (L.), *Terminologie: Approches transdisciplinaires*, Presses de l'Université d'Ottawa. (accepté)
- Dancette (J.), 2007 : « Semantic Relations in the Field of Retailing », dans *Terminology*, n° 13/2, p. 201-223.
- Diderot (D.) et d'Alembert (J.), 1751-1772 : *Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers*, Paris.
- Ertzcheid (O.), 2002 : *Les enjeux cognitifs et stylistiques de l'organisation hypertextuelles: le Lieu, Le Lien, Le Livre*, Université de Toulouse II – Le Mirail, Faculté des Lettres et Sciences Humaines, consulté le 16 novembre 2009, <http://tel.archives-ouvertes.fr/docs/00/04/68/71/PDF/tel-00006260.pdf>.
- Jurica (I.), Mulopoulos (J.) et Yu (E.), 1999 : « Using Ontologies for Knowledge Management: An Information Systems Perspective », dans *Proceedings of Knowledge Creation, Organization and Use*, ASIS Annual Conference, Washington DC, p. 482-496, consulté le 16 novembre 2009, <http://www.cs.toronto.edu/pub/eric/asis99.pdf>.
- Institut universitaire de gériatrie de Montréal, 2008 : *Thésaurus du vieillissement et de la santé*, dernière mise à jour le 16 septembre 2009, <http://www.iugm.qc.ca/thesaurus/>.

- Maroto (N.), 2007 : *Las relaciones conceptuales en la terminología de los productos cerámicos y su formalización mediante un editor de ontologías*, Departament de traducció i comunicació, Universitat Jaume I.
- Maroto (N.), et Alcina (A.), 2009 : « Formal description of conceptual relationships with a view to implementing them in the ontology editor Protégé » dans, *Terminology*, 15/2, p. 232-257.
- Roche (C.), 2003 : « Ontology: a Survey » dans *8th Symposium on Automated Systems Based on Human Skill and Knowledge*. IFAC. September 22-24, 2003, Göteborg, Suède, consulté le 20 novembre 2009, <http://ontology.univ-savoie.fr/condillac/files/docs/articles/2003-ifac-ontology-a-survey.pdf>.
- Varet (G.), 1956 : *Histoire et savoir — Introduction théorique à la bibliographie: les champs articulés de la bibliographie philosophique*, Paris, vol. 12, (Les Belles Lettres), consulté le 16 novembre 2009, ([http://books.google.com/books?id=8ApmLT-DCzUC&printsec=frontcover&hl=fr & source=gbs_v2_summary_r&cad=o#v=onepage&q=&f=false](http://books.google.com/books?id=8ApmLT-DCzUC&printsec=frontcover&hl=fr&source=gbs_v2_summary_r&cad=o#v=onepage&q=&f=false))
- Wikipédia, 2009 : *Contrôle de la qualité des articles*, dernière mise à jour le 14 novembre 2009, [http://fr.wikipedia.org/wiki/Wikip%C3%A9dia#Contr. C3.B4le_de_la_qualit. C3.A9_des_articles](http://fr.wikipedia.org/wiki/Wikip%C3%A9dia#Contr._C3.B4le_de_la_qualit._C3.A9_des_articles).

Annexes

Annexe 1 : Article GOUVERNANCE, extrait de l'*Encyclopédie Wikipédia*

GOUVERNANCE

Chez la plupart de ceux qui, au sein du secteur public comme au sein du secteur privé, emploient le terme de gouvernance, celui-ci désigne avant tout un mouvement de « décentrement » de la prise de décision, avec une multiplication des lieux et des acteurs impliqués dans cette décision. Il renvoie à la mise en place de nouveaux modes de régulation plus souples, fondés sur le partenariat entre différents acteurs.

On distingue deux grands types de gouvernance : la gouvernance d'entreprise pour le secteur privé et la gouvernance politique pour la pensée politique et administrative. En gouvernance politique, on parle de Gouvernance mondiale ou globale, de Gouvernance locale ou Gouvernance régionale en fonction des échelles de gouvernance abordées.

La gouvernance concerne en particulier :

les sociétés par actions, où les actionnaires ne doivent pas être simplement des apporteurs de capital, les institutions publiques, telles que l'État, l'Union européenne, les collectivités locales, pour le respect des droits des citoyens, les organismes sociaux pour le respect des cotisants et bénéficiaires, et les associations pour le respect des membres.

(Wikipédia, Gouvernance, <http://fr.wikipedia.org/wiki/Gouvernance>, dernière mise à jour octobre 2009)

Annexe 2: Article RÉGIME DE PENSION, extrait *du Thésaurus du vieillissement et de la santé* (2003)

| |
|--|
| <p>RÉGIME DE PENSION</p> <p>Anglais: Pension plans</p> <p>Employé pour: Fonds de pension, Plan de pension, Prestation, Rente ø</p> <p>Terme spécifique: <u>Allocation au survivant</u>, <u>Allocation aux anciens combattants</u>, <u>Pension d'invalidité</u>, <u>Pension de sécurité de la vieillesse</u>, <u>Régime de retraite</u></p> <p>Terme associé: <u>Finances personnelles</u>, <u>REER</u></p> <p>Note d'application: Employer comme générique pour tout fonds de pension, ou lorsqu'on parle de cotisants et/ou de cotisations, de prestations et/ou de prestataires. Employer « Régime de retraite » lorsqu'il s'agit d'une pension versée dans le cadre d'un travail.</p> |
|--|

Annexe 3: Article GOVERNANCE, extrait de *ILO Thesaurus*

| |
|--|
| <p>GOVERNANCE</p> <p>FRE: gouvernance SPA: gobernabilidad</p> <p>Scope Note: The manner in which power and authority are exercised by both public and private bodies. Covers management, legal framework, accountability and transparency.</p> <p>Related Terms: government, legal system, management, public administration, responsibility.</p> |
|--|

Annexe 4: Article GOUVERNANCE D'ENTREPRISE, extrait du DAMT. (Nous ne reproduisons pas les tableaux de relations puisque nous les avons présentés dans le texte.)

GOUVERNANCE D'ENTREPRISE

Définition: L'organisation du pouvoir au sein d'une société ou d'une entreprise visant à un meilleur équilibre entre les instances de direction, les instances de contrôle et les actionnaires ou sociétaires.

Description: Le concept de gouvernance d'entreprise est né de la nécessité d'examiner de façon critique les mécanismes qui encadrent la prise de décision dans les entreprises et de l'exigence, de la part de la société civile, d'une plus grande transparence quant aux activités des entreprises et à leur responsabilité sociale. « Une bonne gouvernance d'entreprise atteindra son objectif en établissant un équilibre adéquat entre l'esprit d'entreprise et le contrôle, ainsi qu'entre la performance et la conformité à ces règles. » (Le code belge de gouvernance d'entreprise, 2004, consulté le 10-06-2009)

Dans le contexte actuel, ce concept est assimilé à celui de gouvernance mondiale. « Pour la Commission de la gouvernance globale, située dans l'orbite des Nations unies, la gouvernance est la somme des pratiques par lesquelles individus et institutions, tant privées que publiques, gèrent leurs affaires communes. C'est un processus continu, par lequel les conflits d'intérêts peuvent être surmontés et une démarche coopérative mise en œuvre. » (Moreau-Defarges, P., G7-G8 et gouvernance économique mondiale, Annuaire français des Relations internationales, 2000) Droit flexible et gouvernance

La gouvernance d'entreprise, tout comme les codes de conduite ou codes de bonnes pratiques, relève d'un droit défini hors de tout processus législatif, dépourvu de sanctions publiques, à savoir un droit non contraignant, un droit « mou » ou « flexible ».

Annexe 5 : Page d'accueil du DAMT

Analytical Dictionary of Globalization and Work
Dictionnaire analytique de la mondialisation du travail - Diccionario analítico de la globalización del trabajo
 Comments via the forum: [The DAMT](#)

Welcome to the Online Thesaurus System

Jeanne Dancette
 Département de linguistique et de traduction, Université de Montréal

The dictionary includes some 5000 terms. The user can find:

- Translations
- Simple definitions
- Encyclopedic descriptions
- Clusters of related terms
 - Official websites
 - Web resources

To find a term in either English, French or Spanish, click on one of the letters to the left.

[Jeanne Dancette Web page](#) [CRIMT Web page](#)

[Click here for more information](#)

Index

Alphabetical index

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [R](#) [S](#)
[T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

Index of terms by domain

[Click here to open the index](#)

Index

Alphabetical index

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [R](#) [S](#)
[T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

Index of terms by domain

[Click here to open the index](#)

Annexe 6 : Liste de relations étiquetées utilisées dans le DAMT

Sept classes de relations sémantiques :

Equi: équivalences interlinguistiques

Gen/Spec: relations de générique à spécifique

Part/Tot: partie-tout

Contr: termes qui s'opposent par au moins un trait

Sing/Mult: relations de singulatif à collectif

Verb/nom: verbe dérivé d'un nom ou entrant en collocation avec un terme

Adj/nom: adjectif dérivé d'un nom

Sept classes de relations conceptuelles :

| |
|---|
| <p><u>Act</u>: personne ou organisation ou groupe d'individus qui fait l'action ou contribue de façon typique au processus → Action : chose, ou processus fait par Actor</p> <p><u>Fact-Result</u>: processus, système, philosophie qui contribue à un état ou processus en résultant → Résultat : ce qui résulte de Facteur</p> <p><u>Instr-Real</u>: activité ou document permettant d'établir ou de créer qqch, ou d'établir un droit pour qqn → Réalisation : ce qui est rendu possible grâce à Instrument</p> <p><u>Leg</u>: législation (norme, règle, loi, accord, code, déclaration) ayant effet sur qqch ou qqn → Légiféré : entité affectée par le document</p> <p><u>Loc</u>: entité physique ou abstraite (usine, pays, zone, secteur) dans laquelle se trouve l'entité localisée (chose, phénomène, activité, personne) → Localisé : l'entité située dans Locus</p> <p><u>Prop</u>: propriété, qualité ou attribut caractérisant une personne physique ou morale, ou un groupe de personnes, une industrie ou un pays → Prop : la personne ou le groupe qui détient la propriété</p> <p><u>Quant</u>: quantification ou mesure rattachée à un phénomène, une activité ou une entité → Quant : entité quantifiée</p> |
|---|

Une classe de « Concepts proches » (concepts auxquels il est habituel de référer en rapport avec le sujet discuté dans l'article).

Pierre Lerat,
Équipe Condillac,
Université de Savoie,
Le Bourget du Lac,
France,
pierre.lerat@wanadoo.fr.

Communautés de travail et diversité terminologique : le préemballage, ses conceptualisations et ses enjeux

Résumé

On part d'objets (les préemballages), on décrit leurs conceptualisations partagées, on en tire des champs conceptuels, on finit par des dénominations situées. Le fondement linguistique est le dernier Harris. Par exemple, *préemballage* peut être défini comme une opération sur un article ou un produit, mais le même nom désignant un objet concret (contenant et contenu) n'est pas prédicatif. Le champ conceptuel de < préemballage > partage avec les ontologies des relations prédicatives classiques. Pour tenir compte de la polysémie et de la synonymie en traduction et documentation, il faut ajouter deux prédicats métalinguistiques : « se dit en » et « se dit aussi ».

Mots-clés : terminologie ; ontologie ; concept ; prédicat ; information.

Introduction

Mon propos est l'élaboration d'un champ conceptuel du mot français *préemballage* et du mot portugais *pré-embalagem*. Il s'agit d'illustrer par l'exemple une méthode de travail complémentaire de celle d'une équipe d'informaticiens qui développe des systèmes conceptuels en vue d'applications, notamment des mémoires d'entreprises. Mon étude, un peu réorientée par rapport au résumé soumis, distingue typographiquement les types de relations et d'entités traitées (<Concepts >, «dénominations » et objets).

1 Les domaines

À quel domaine appartient le préemballage ? D'après les thésaurus, l'agro-alimentaire est concerné au premier chef, mais aussi le commerce, la pharmacie, etc. On verra plus loin des références à l'OIML (Organisation internationale de métrologie légale) et à l'OIV (Organisation internationale de la vigne et du vin), et aussi à *Google*, pour les sites de fabricants d'emballages. Bref, beaucoup de métiers sont concernés.

2 Les communautés de travail

Les domaines restent des repères documentaires utiles, mais une terminologie est surtout propre à une communauté de travail. De travail, plutôt que de pratique, car le concept de communauté de pratique est surtout utilisé en pédagogie, alors qu'ici ce qui importe est l'ensemble des acteurs professionnels, y compris les experts que consultent la FAO et l'UE avant d'édicter leurs normes.

Les communautés de travail sont ancrées dans l'histoire, comme le soulignent cruellement les suppressions d'emplois. En gros, en 1974 l'UE se saisit des enjeux des préemballages en vue de régler la libre circulation des biens ; en 1985 la FAO publie une norme à ce sujet ; en 1997 l'OIML s'attache principalement aux matériaux des préemballages ; en 2002 l'UE réexamine le dossier sous l'angle de la sécurité des aliments ; en 2006 l'oiv associe étiquetage et préemballage ; en 2009 la FAO s'inspire de l'OIML.

Une terminologie normalisée à l'échelle mondiale est une terminologie qui fait l'objet de définitions et de dénominations consensuelles dans toutes les communautés de travail concernées. Celles-ci vont des instances officielles mondiales, continentales, nationales et par branches jusqu'au jargon d'entreprise.

3 L'émergence des préemballages

L'apparition du mot *préemballage* est un épiphénomène dont la datation importe peu : 1983 selon le *TLFi*, 1975 selon le *GDT*, 1974 au moins selon l'UE, certainement plus tôt chez les professionnels. Ce qui importe, c'est que le phénomène, et non pas le mot, bénéficie massivement de l'usage récent de plastiques nouveaux, à la fois solides, souples et pas chers (notamment les polyéthylènes basse densité).

4 La conceptualisation des objets

Nous avons tous en commun des concepts comme le but ou la cause, qui appartiennent à « notre ontologie naturelle partagée » (Gross 2004 : 12). Tout autre est le mode d'existence des concepts techniques, qui sont la matière de la terminologie depuis sa mondialisation au sein de la CEI (Commission électrotechnique internationale, 1906) puis à l'ISO. Il s'agit d'artefacts, qui sont des objets concrets, comme les aspirateurs, ou abstraits, comme le droit de rétractation du consommateur dans la vente en ligne. Leurs conceptualisations sont nécessairement consensuelles, d'où la nécessité de comités techniques multiples au sein de l'ISO, par exemple. Il ne s'agit pas de catégorisations par un « je-ici-maintenant » mais, pour tel objet, de connaissances partagées par les professionnels : la classe dont il relève, son genre prochain (immédiatement supérieur) et ce que le TC 37 de l'ISO appelle ses « caractères », ce qui correspond aux « propriétés communes à l'ensemble possible des référents » (Harris 2007 : 65).

Il faut ensuite s'entendre sur les dénominations. Objets, concepts, dénominations... un air de déjà vu. Effectivement, la terminologie à base d'objets conceptualisés implique le triangle sémiotique cher à Frege, qui remonte au moins à l'Antiquité.

5 Le champ conceptuel

Ce qui rend possible et fécond le travail commun de terminologues et d'ontologues, c'est que les premiers aboutissent à un champ conceptuel et que les seconds en partent (Roche 2008). Voici les principales relations constitutives du champ conceptuel de < préemballage >.

5.1 Le genre prochain

En 1974, la définition officielle de *préemballage* est « ensemble d'un produit et de l'emballage individuel dans lequel il est préemballé » (UE 1974). Le définisseur lexicographique « ensemble » est trop général pour constituer un « genre prochain » plausible. Pour les professionnels, c'est une « unité de vente ».

En 2008, nouvelle définition : « Emballage de vente, etc. ». S'agirait-il seulement du contenant ? Non, car « emballage de vente » est défini comme « partie d'un lot et de son contenu conditionnée individuellement de manière à constituer, au point de vente, une unité de vente destinée à l'utilisateur final ou au consommateur ». Ce qui nous ramène à « unité de vente », dont le préemballage est un cas particulier : unité de vente + différence spécifique.

5.2 La relation *partie de*

Cette relation entre concepts incluant et inclus ne permet pas l'héritage des propriétés. Ainsi, le préemballé est une partie du préemballage au sens de l'« ensemble d'un produit et de l'emballage individuel dans lequel il est préemballé » (UE 1974), l'autre partie étant le matériau d'emballage, mais ce qui est vrai du préemballage ainsi conçu ne l'est pas forcément des préemballés.

Ex. : l'étanchéité d'un préemballage dépend entièrement des matériaux d'emballage, non du contenu.

La relation est donc orientée :

< Préemballage > -- a pour partie -- → -- < produit > & < emballage individuel >

5.3 La relation *fr* (pour « se dit de préférence en français »)

Alors que le choix du nom du concept peut être distinct à la fois des termes normalisés et des termes d'usage, puisqu'il s'agit de gestion interne du champ conceptuel, il est souhaitable que le traducteur soit orienté vers la dénomination la plus recommandable. Entre le verbe et sa nominalisation, il vaut mieux choisir la nominalisation, pour des raisons de commodité et de fréquence qui ont été présentées ailleurs (Lerat 2007 : 83), même si dans les discours des professionnels, à l'oral, les

formes verbales sont plus courantes (voir la communication de Dardo de Vecchi); au reste, *préconditionnement* est plus fréquent que *préconditionner* sur le Web. D'où la relation orientée < Préconditionnement > ---- fr -- → -- « preconditionnement »

5.4 La relation *frs* (pour « synonyme acceptable en français »)

Les langues spécialisées n'échappent pas à la synonymie, pas plus qu'à la polysémie. Ainsi, le terme normé *préconditionnement* est en concurrence avec l'emploi prédicatif (insaturé) de *préemballage* (« opération qui consiste à emballer des produits », selon le *Guide terminologique* du CILF), homonyme de l'objet dit également *préemballage*. C'est cet usage courant que signale ici *frs*.

Ex. : < Préconditionnement > ---- frs -- → -- « préemballage »

6 Une description minimale

J'ai présenté ailleurs (Lerat 2009a et Lerat 2009b) les principales caractéristiques de la « description minimale » (Harris 2007 : 30) appliquée par Harris aux « *subject-matter sublanguages* » (Harris 1988 : 37), ce que j'appelle « les langues spécialisées ». Il s'agit de généralisations tirées des discours de spécialités, tout comme une langue totale (« *whole language* », Harris 1988 : 104) est une généralisation à partir des usages d'un idiome. Je résume ici l'essentiel.

6.1 Conception extralinguistique du sens

« Le sens que prend un mot est manifestement son association avec certains objets ou états du monde qu'on appelle ses référents » (Harris 2007 : 64). On reconnaît ici l'influence de deux grands linguistes américains: Sapir et Bloomfield.

6.2 Conception discursive de l'information

L'information au fil du discours est principalement une interprétation des enchaînements. Les langues spécialisées n'échappent pas à l'énonciation, et une anaphore dans un texte d'immunologie est toujours à interpréter par un sujet humain, qui doit « reconstituer l'antécédent par inférence, du mieux qu'il peut » (Harris 2007 : 73). Ce qui, en revanche, réduit à néant le risque de mauvaise interprétation d'un mot spécialisé polysémique, c'est que « les mots requis sont identifiés par ce qu'ils requièrent en retour » (Harris 2007 : 36).

Ex. 1: préemballage de l'eau (*Google*) désambiguïse le sens de préemballage et sélectionne l'interprétation « preconditionnement » :

< Préconditionnement > ---- a pour argument -- → -- < Eau >

Ex. 2: préemballage contenant un produit (OIML 1997) exclut la valeur prédicative :

< Contenir > ---- a pour argument -- → -- < Préemballage > & < Produit >

Ce sont des collocations si l'on veut, mais des « collocations conceptuelles » (Lerat 1995 : 104), c'est-à-dire les arguments typiques d'un prédicat donné dans tel discours spécialisé, et non pas de simples cooccurrences fréquentes.

6.3 Saturation et insaturation

La relation *a pour argument*, comme les autres, est orientée : elle part du mot qui a besoin d'un ou plusieurs arguments.

Ex. : le préconditionnement est le préconditionnement de quelque chose, qui ne peut être qu'un article (emballage interne compris) ou un produit (contenu dans un matériau de préemballage), alors que la liste des prédicats possibles pour < Produit >, comme celle des arguments possibles pour < Contenir >, est sans limite calculable.

7 Les enjeux du préemballage

Une terminologie à orientation ontologique ne décrit pas pour décrire, mais pour agir sur les choses grâce à une meilleure connaissance des réalités. D'où l'importance de la fonction dans une ontoterminologie (Calberg 2009).

Ex. : < préemballage > a pour fonction < sécurité des aliments > (point de vue médical : ue 2002) & < sécurité sanitaire > (point de vue politique : OMS 2004), ce qui est résumé sur le schéma par :

< Préemballage > ---- a pour fonction -- → -- < Sécurité >

L'UE s'efforce de prévenir les risques juridiques, commerciaux et sanitaires. Le préemballé pose des problèmes sérieux de date de péremption, de dosage des conservateurs, etc. L'autre partie du préemballage, les matériaux d'emballage, est du ressort de la métrologie légale : poids brut et poids net, résistance des matériaux, tels sont les attributs qui l'intéressent avant tout. La FAO s'inquiète à juste titre des conditions d'acheminement, mais l'OMS voit dans le bon usage des préemballages plus de raisons d'espérer que de craindre.

Ex. : une « cure antipaludique complète sous enveloppe thermofermée dans un emballage primaire d'aluminium/PVC » (OMS 2004).

Conclusion

Le champ conceptuel proposé est à base de linguistique générale non saussurienne. Affaire à suivre, donc. Après tout, la dénomination n'est que, pour reprendre les termes de Harris, « l'institutionnalisation de l'habitude, de ce qui est commode et de ce qui fait sens » (Harris 2007 : 91) ; bref, une affaire de « mémoire des mots », au sens des sciences cognitives.

Si le travail terminologique est bien fait, il rend possible une systématisation logique qui le valide par des contrôles de cohérence.

Ex. : puisque les définitions de « préemballage » et de « pré-emballagem » (même genre prochain et même différence spécifique) concordent, les liens suivants sont valides :

- < Préemballage > ---- *fr* -- → -- « préemballage »
- < Préemballage > ---- *pt* -- → -- « pré-emballagem »

Cela permet d'inférer l'équivalence de *préemballage* et de *pré-emballagem*.

Il existe divers outils d'aide à la traduction, mais leur prolifération n'empêche pas les traducteurs spécialisés de « demander instamment une gestion responsable de la variation linguistique » (Maldussi 2008 : 90). Espérons donc qu'une aide conceptuelle à la traduction leur offrira un peu plus de sécurité. Il faut pour cela une définition descriptive cumulative, un jeu prédéfini de relations dans un champ conceptuel, et si possible, dans le cadre d'une ontologie, un concept spécifié logiquement.

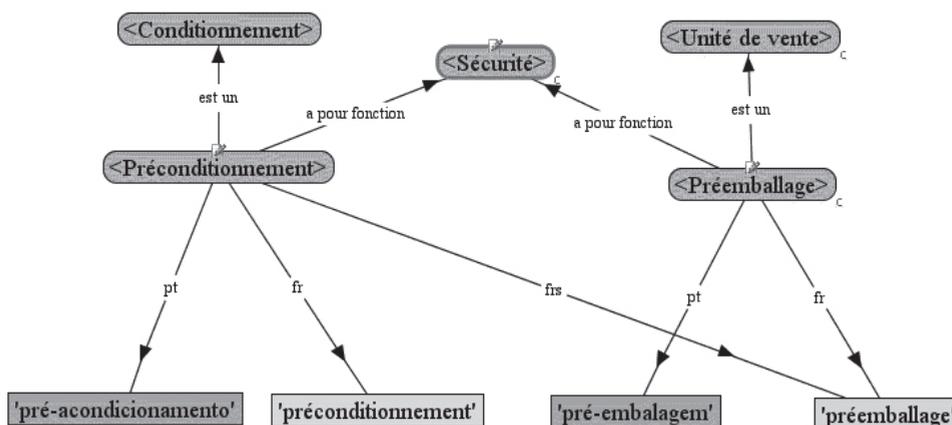
Bibliographie

- Base terminologique*: Conseil international de la langue française, consultée le 20 septembre 2009, www.cilf.org.
- Calberg (M.), Lerat (P.) et Roche (C.), 2009 : « Quelle place accorder aux corpus dans la construction d'une terminologie ? », à paraître dans les actes de *TOTH 2009*, Annecy, Institut Porphyre, www.porphyre.org.
- GTD*, 2006 : *Le grand dictionnaire terminologique*, Québec, Office québécois de la langue française, consulté le 20 septembre 2009, www.granddictionnaire.com.
- Gross (G.) et Prandi (M.), 2004 : *La finalité. Fondements conceptuels et genèse linguistique*, Bruxelles-Louvain-la-Neuve, De Boeck-Duculot.
- Harris (Z.), 1988 : *Language and Information*, New York, Columbia University Press; trad. fr. *La langue et l'information*, Ibrahim (A.H.) et Martinot (C.) éd., Paris, Cellule de recherche en linguistique, <http://crl.exen.fr>.
- Lerat (P.), 1995 : *Les langues spécialisées*, Paris, PUF.
- Lerat (P.), 2007 : « Les nominalisations en *tion* dans un texte techno-administratif » dans *TOTH 2007*, Annecy, Institut Porphyre, p. 53-72, www.porphyre.org.
- Lerat (P.), 2009a : « La combinatoire des termes. Exemple : nectar de fruits », dans *Hermès*, 42-2009, p. 211-232.
- Lerat (P.), 2009b : « Variabilité et harmonisation terminologique », communication au *Convegno Assiterm 2009*, à paraître dans les actes, www.publiforum.farum.it.
- Maldussi (D.), 2008 : *La terminologia alla prova della traduzione specializzata*, Bologna, Bonomia University Press.
- Roche (C.), 2008 : « Faut-il revisiter les principes terminologiques ? » dans *TOTH 2008*, Annecy, Institut Porphyre, p. 53-72, www.porphyre.org.
- TLFi*: *Trésor de la langue française informatisé*, consulté le 19 septembre 2009, <http://atilf.atilf.fr>.

Documents cités

- 1974 : Directive 75/116/CEE du Conseil, du 19 décembre 1974, concernant le rapprochement des législations des États membres relatives au préconditionnement en volume de certains liquides en préemballages.
- 1985 : Norme générale CODEX pour l'étiquetage des denrées alimentaires préemballées, www.codexalimentarius.com.
- 1997 : Exigences pour l'étiquetage des produits préemballés, OIML E 79, www.OIML.org.
- 2002 : Règlement (CE) n° 178/2002 [...] fixant des procédures relatives à la sécurité des denrées alimentaires.
- 2004 : « Sources et prix d'une sélection de produits pour la prévention, le diagnostic et le traitement du paludisme », OMS, www.who.int.
- 2006 : Harmonisation de l'étiquetage des vins. Responsable des emballages, résolution ECO 6/2006, www.oiv.com.
- 2008 : Règlement (CE) n° 1221/2008 [...] en ce qui concerne les normes de commercialisation.
- 2009 : Commission du Codex alimentarius, CX/FL 09/37/12, ftp://ftp.fao.oeg/codex/ccfl37/fl37_12f.pdf.

Annexe



Champ conceptuel de < préemballage >

*Taoufik Massoussi,
Gérard Petit,*
LDI (UMR 7187) CNRS et Université de Paris 13,
Villetaneuse, Boulogne-Billancourt,
France,
taoufik.massoussi@viavoo.com,
gerardpetit@wanadoo.fr.

***Métalangue*, une base de données multilingue**

Résumé

Nous présentons dans cet article une base de données des termes de la métalangue, développée au LDI (UMR 7187), et destinée à jouer plusieurs rôles : (i) répondre à un besoin urgent de standardisation conceptuelle et de cohérence terminologique, surtout dans les travaux contrastifs ; (ii) mettre à la disposition de la communauté un outil d'évaluation interne permettant de mesurer les flottements des emplois des termes, et (iii) fournir une nomenclature multilingue, issue de la traduction d'ouvrages de linguistique, laquelle donnera lieu à des descriptions de la métalangue dans le cadre du traitement automatique des discours spécialisés.

Mots-clés : métalangue ; base de données ; observatoire ; corpus ; extraction.

Introduction

La métalangue utilisée par les linguistes et les grammairiens est le lieu d'un paradoxe. D'une part, elle n'est pas unifiée, et ce, depuis l'origine de la réflexion sur le langage et les langues. La raison en tient pour partie à son positionnement interdisciplinaire, entre linguistique, grammaire, lexicographie¹, philosophie, sociologie, psychologie, logique, informatique, pédagogie ou encore politique. Par ailleurs, si d'importantes zones de recouvrement existent entre ces aires, aucun consensus autre qu'approximatif et minimal n'a pu émerger à travers les siècles, encore moins depuis l'avènement de la linguistique comme science du langage. Chercher une issue normative n'aurait guère de sens aujourd'hui. Il est en revanche urgent de proposer au linguiste et au grammairien un observatoire qui permettrait au chercheur et au traducteur d'optimiser leurs productions. C'est la fonction de la base de données *Métalangue*.

Une première partie de cette étude exposera la problématique qui a présidé à l'établissement de cette base, les objectifs assignés à celle-ci en même temps que les moyens consacrés à son développement. Dans un second temps, une description

¹ Bien qu'intégrées aujourd'hui à la linguistique, la grammaire et la lexicographie la précèdent historiquement.

générale de la base sera proposée. Une troisième partie enfin sera consacrée à la gestion des corpus, en phase avec les objectifs précédemment définis.

1 La métalangue : le linguistique et la gestion du babélisme

1.1 Un déficit de structurel de standardisation

Toute discipline mesure sa vitalité au nombre de ses publications, à la qualité des débats qui la traversent et à la conscience métalinguistique qu'elle développe sur elle-même. La linguistique, en sa qualité de science du langage, n'échappe pas à la règle, avec toutefois une particularité : bien qu'assimilable à une terminologie, sa métalangue n'a pas connu, contrairement à d'autres disciplines, la stabilisation nécessaire à l'univocité de la référencement et des échanges.

Il en résulte une situation doublement dysfonctionnelle. D'abord en ce qu'une même entité peut se voir indexée par des expressions extrêmement variées d'une branche à l'autre, voire entre écoles et auteurs. La diversité des différents modes de désignation des séquences polylexicales est connue. Nous l'illustrerons à l'aide d'exemples récents empruntés respectivement à Marie-Françoise Mertureux (2008) et Gaston Gross (1996) [nous soulignons par l'italique les expressions communes aux deux auteurs] :

- *mot composé*, lexème composé, expression figée, *séquence figée*, signe polylexical, *construction figée*, syntagme, structure syntaxique, lexème composé, signe codé, suite lexicalisée, lexie, synapsie, expression, expression figurée, expression idiomatique, formule, idiome, *locution*, séquence, séquence signifiante, *suite*.
- *mot composé*, mot construit, mot simple, catégorie simple, composé complexe, composé endocentrique, composé exocentrique, suite composée, terme composé, composé sur thèmes savants, catégorie polylexicale, *construction figée*, élément polylexical, locution figée, mot polylexical, *séquence figée*, suite affectée par le figement, unité polylexicale, unité figée, suite figée, suite métaphorique figée, suite sémantiquement figée, construction générée par la syntaxe, construction régulière, composé à élément tête, suite contrainte, suite contrainte lexicalement, locution contrainte, suite à actualisation contrainte, suite à actualisation contrainte, suite à élément non actualisé, suite exocentrique opaque, suite inanalysable, suite opaque, suite sémantiquement figée, suite métaphorique figée, *locution*, séquence proverbiale, *suite*, suite « gelée », unité fonctionnelle, unité intermédiaire.

L'effet de dispersion, déjà considérable, ne sera qu'amplifié si l'on étend la comparaison à d'autres auteurs, spécialistes ou non de la polylexicalité, ou même à une autre thématique. Il en résulte un effet de babélisme exacerbé : *babélisme* du fait que chaque école et chaque auteur tend à se comporter de manière autonome, utilisant son propre réseau d'appellations, coupées à la fois de son lecteur ou des autres chercheurs ; *exacerbé* dans la mesure où, à l'exception des chercheurs à l'origine

d'une conceptualisation systématique et strictement contrôlée, ce(s) réseau(x) n'est (ne sont) pas stabilisé(s) chez tel auteur mais en constante évolution.

Phénomène inverse la même appellation peut renvoyer à des entités différentes d'un auteur à l'autre : *lexème* chez Mortureux vs chez Polguère ou Mel'čuk, *lexie* chez Barthes vs chez Polguère/Mel'čuk, *mot possible* chez Corbin vs chez Temple, *discours* chez Harris vs chez Benveniste, *notion* et *concept* chez Gentilhomme vs chez Culioli ou Grize, etc. Les exemples pourraient être multipliés à l'envi. Le constat serait identique : loin d'être constituée en système cohérent la métalangue se comporte de manière erratique. Lui fait défaut la propriété qui consignerait son statut terminologique et lui conférerait en même temps une reconnaissance et une pérennité : la standardisation. Que celle-ci soit pensée sous le mode de l'autorégulation (la « normaison » au sens de Gaudin 1993) ou de l'intervention institutionnelle (la normalisation), la standardisation est toujours demeurée une préoccupation résiduelle chez les linguistes et grammairiens. Pour cette raison, et compte tenu de la tendance observable à longue échéance, une attitude normative ne ferait qu'ajouter à la confusion d'ensemble.

1.2 L'effet boule de neige

La situation décrite à l'alinéa précédent ne peut être tenue comme une simple particularité fâcheuse ou cocasse, et ce, pour au moins deux raisons : d'abord parce qu'elle impacte négativement la recherche et sa diffusion ; ensuite parce qu'elle hypothèque l'équipement métalinguistique des langues.

Les jeunes chercheurs, mais aussi des auteurs chevronnés, utilisent des « termes » et des « concepts » métalinguistiques hérités d'une tradition ou de transferts. L'observation démontre que ces auteurs ne maîtrisent pas intégralement les unités métalinguistiques auxquelles ils ont recours : le déficit porte soit sur la dénotation, soit, plus fréquemment, sur les prérequis théoriques attachés à tel « terme » introduit dans la métalangue au cours d'un transfert, soit du fait de son appartenance au champ spécifique de tel courant ou école. Ainsi, le terme de *dénomination*, tel qu'il est repris dans le domaine de la morphologie constructionnelle (modèle associatif stratifié de Corbin), est emprunté *in extenso* à un article de Georges Kleiber (1984), lequel l'avait emprunté à la logique (entre autres Stuart Mill). Les difficultés d'interprétation que posait la nécessaire condensation de l'article de 1984 n'ont pas été suffisamment prises en compte par les morphologues, pour qui la dénomination, si elle constitue un concept central, n'a jamais été supposée faire l'objet d'investigations conséquentes. Pareillement, le modèle d'analyse sémantique procédant par propriétés intrinsèques et extrinsèques (Cadiot *et al.*) a construit la dénomination comme le contrepoint de son appareil théorique. Toutefois, c'est en se fondant sur une approche stéréotypique et non conceptualisée du phénomène (Petit 2001). Enfin, la différence d'acception entre les valeurs respectives de *dénomination* en linguistique et celles qui sont les siennes dans le domaine de la terminologie ne se recouvrent que très partiellement (Humbley 2001). Prenons un autre exemple : il n'est pas rare de rencontrer dans des travaux d'étudiants de Master ou

de doctorants des contradictions liées à la cooccurrence de termes appartenant à des appareils théoriques incompatibles. Ainsi lit-on parfois que des « règles de construction » s'appliquent à des « morphèmes ». Or, une petite investigation dans le domaine des morphologies par règles montre que celles-ci sont par principe... amorphématiques, donc récusent la notion de morphème.

De fait, le chercheur qui manipule la métalangue (en réception ou en production) doit être informé à la fois de l'histoire théorique des termes métalinguistiques et des prérequis annexés à leur intégration. Ainsi seraient évitées, en aval, des productions présentant des déficits ponctuels ou structurels de conceptualisation. Plus fondamentalement seraient écartées des confusions dommageables à la cohésion de la réflexion.

En matière de traduction, la question se pose de manière plus aiguë du fait de la correspondance à établir entre métalangue source (MTS) et métalangue cible (MTC). Deux configurations s'observent. Des langues présentent des notions homonymes dans leurs dénominations : ainsi, le grec moderne et le français connaissent tous deux la notion de « futur simple ». Toutefois, dans l'une et l'autre langues, celle-ci dénote des entités différentes (temps non composé/aspect non accompli pour le français ; temps non composé/aspects non accompli et ponctuel² pour le grec). De même, la notion de « racine » ne recouvre pas en arabe et en français la même réalité. Ceci est dû au fait que chaque langue a un système de dérivation fondamentalement différent. Si la dérivation en français est syllabique, en arabe elle se fait par schèmes. En français, « racine » renvoie à un « élément de formation commun aux unités lexicales d'une même famille » (Neveu 2004). On dira par exemple que la racine *respect-* permet de dériver les mots *respectueux*, *respecter*, *respectable*, entre lesquels il y a un invariant sémantique. La dérivation se fait donc par un jeu d'affixes. S'agissant de l'arabe, le terme *racine* désigne un schème trilitère, ou quadrilitère. Ce schème correspond à une matière consonantique (par exemple *كتب ktb* [ce qui se rapporte à l'écrit] ; *رسم rsm* [ce qui se rapporte au dessin], etc.), versée dans des moules prédicatifs (agent, action) ou argumentaux (lieu, temps, etc.).

Seconde configuration : des notions sont absentes en MTS ou en MTC, car elles indexent des entités inexistantes pour la langue concernée : « infinitif » et « nominal » en grec moderne, « subjonctif » en arabe, « médiopassif » en français. Leur traduction ne peut procéder que de l'approximation ou de la périphrase.

Enfin, la gestion du babélisme d'une MTS ou d'une MTC (voire de deux), introduit une perturbation qui implique des choix conscients concernant les diverses variables d'une même donnée à traduire depuis une MTS (*cf.* plus haut Gross et Mortureux), compte tenu de celles qui sont disponibles et pertinentes en MTC. La configuration où une correspondance biunivoque existerait entre les diverses variables d'une même donnée en MTS et MTC reste utopique. Le non-isomorphisme des métalangues respectives implique des péréquations qui, si elles doivent

2 En grec, le contenu véhiculé par « simple » ne s'oppose pas à composé, mais à « continu ».

rendre compte des *denotata*, n'en restent pas moins tributaires des effets de sens à translater. À cela s'ajoute le poids de traditions qui, comme c'est le cas pour l'arabe, ont marqué les pratiques de traduction. Elles ont imposé des normes qui constituent autant de points d'achoppement d'une traduction privilégiant l'adéquation à MTS.

2 La base *Métalangue*

2.1 Un observatoire et un outil

La base *Métalangue* est développée au LDI dans le but de remédier aux difficultés précédemment évoquées. Créée en 2008, elle comporte actuellement 7 000 fiches. Cette base est conçue comme un observatoire, c'est-à-dire dans une visée descriptive. Dans le courant de l'année 2010, elle sera accessible sur internet moyennant une inscription et un mot de passe. La première fonction de la base est épistémologique. Il s'agit de :

- recenser et analyser les termes spécifiques à la métalangue ;
- dresser une cartographie conceptuelle et dénomminative du domaine (inventaire des termes employés et statut sémiotique) ;
- localiser et mesurer l'évolution subie par le vocabulaire des sciences du langage (*sème* ne désigne plus aujourd'hui un trait nécessairement pertinent) ;
- identifier les termes selon qu'ils sont créés, empruntés, banalisés. Les auteurs ne signalent pas toujours les emprunts qu'ils font, ceux-ci se diluant dans la masse des unités banalisées de la métalangue. Les dénominations forgées ne sont pas systématiquement assorties du balisage textuel (typographique, définitionnel) nécessaire à leur identification. Elles sont par ailleurs susceptibles d'évoluer dans la pensée d'un même auteur (p. ex. la notion de « propriété extrinsèque » au fil des publications de P. Cadiot, celle de « mot possible » chez D. Corbin et sa reprise chez M. Temple) ;
- produire une analyse et une représentation de la variation subie par les termes métalinguistiques (voir plus bas § 2.2.) ;
- observer la relation signalétique d'un terme relativement au paradigme théorique à l'intérieur duquel il est né et/ou a acquis sa valeur d'emploi. Certains termes sont symptomatiques des théories qui les ont forgés et restent marqués par cette mémoire discursive : « propriété extrinsèque » (Cadiot *et al.*), « élément de nomination » (Boisson), « sens prédictible » (Modèle associatif stratifié), « lexie » (Modèle Sens-Texte), « mot » (Benveniste).

En second lieu, elle doit constituer un outil d'information et de formation destiné aux (jeunes) chercheurs. Actuellement, disposer d'une traçabilité des termes est un impératif si l'on veut éviter une appropriation et un usage incontrôlés car insuffisamment informés des termes de la métalangue. La base de données dispose de champs spécifiques (*cf.* plus bas § 2.2. : <concept>, <statut>, <définition>, <notes>) destinés à établir, dans toute la mesure du possible, le positionnement sémiotique et

sémantique de tel terme chez tel auteur, et les prérequis qui y sont attachés compte tenu de son évolution théorique. Comme toute production spécialisée, les textes (méta) linguistiques sont traversés d'interdiscours plus ou moins conscientisés. L'objectif n'est pas de fournir en sortie un standard en fonction duquel affecter des coefficients à des textes reçus ou produits, mais des modèles de formulations disponibles permettant au chercheur de lire les zones de fractures qui traversent une production et de mesurer le degré de contrôle par l'auteur des concepts qu'il convoque. Le chercheur pourrait alors assurer davantage la cohérence conceptuelle de son propos (éviter les approximations, faux sens et autres interférences, etc.)³.

Le troisième objectif de la base *Métalangue* est de constituer une aide à la traduction et à l'équipement métalinguistique. Chaque fiche est dotée de deux champs : <traduction> et <notes traduction>. Actuellement la traduction s'effectue vers l'arabe.

Dans le cadre de l'action AUF-LTT PA.2091RR803, une équipe de recherche spécifique a été constituée en Tunisie. Elle prend en charge la traduction vers l'arabe des fiches déjà recensées, mais aussi la construction d'une base arabe, isomorphe et couplée à la version française par des liens hypertextes. Ceux-ci permettent au chercheur et au traducteur de passer de l'une à l'autre pour compléter son information dans MTC ou MTS. L'objet d'une interface de traduction est également de procéder à un équipement métalinguistique suivant l'axe Nord-Sud. Toutefois, l'équipement n'est pas pensé comme orienté dans une direction unique, mais doit profiter également à l'équipement de la langue source. Dans une langue comme dans l'autre, l'équipement n'est pas conçu pour déboucher sur la production d'une norme.

2.2 Structure de la base

Différents types de champs permettent de porter des informations syntactico-sémantiques (catégorie grammaticale, fonction primaire), d'autres ont trait aux éléments du paratexte (contexte, termes apparentés, termes opposés), d'autres ont trait au caractère multilingue (traduction), au corpus (source, auteur) ou encore au travail de suivi au sein de l'équipe (contributeur, date de mise à jour, etc.). Nous développons successivement les champs syntactico-sémantiques, les champs du paratexte et les champs contrastifs.

2.2.1 Champs syntactico-sémantiques

Il s'agit d'indiquer la catégorie grammaticale, la fonction primaire, le (ou les) domaine(s) au (x) quel(s) le terme est rattaché, et le statut du concept véhiculé par ce terme. Ces champs appellent deux remarques : la première a trait au traitement automatique de la métalangue, et la deuxième à l'interaction entre les différents champs de la base.

³ Lors de la séance de questions, un chercheur en TAL a souligné la nécessité pour la communauté de disposer d'outils permettant à chacun de se familiariser avec les concepts des disciplines connexes.

Les deux premiers champs, catégorie grammaticale et fonction primaire, fournissent une catégorisation fonctionnelle des termes de la base. En effet, si la catégorie grammaticale est une information relativement classique, elle devient plus opératoire lorsqu'elle est couplée à l'indication de la fonction primaire du terme en question. Autrement dit, le fait d'indiquer que *dénomination* est un substantif n'est plus banal lorsqu'il est complété par l'indication qu'il s'agit tantôt d'un prédicat (acte de dénommer), tantôt d'un argument (le résultat de l'acte de dénommer). Cela permet de traiter des emplois polysémiques internes au domaine des sciences du langage, en consacrant une double entrée à ces termes⁴. Actuellement, un axe de recherche au LDI consiste à étudier les sciences du langage comme un discours spécialisé, que l'on traite au moyen des mêmes outils mis en place pour la description de la langue générale et d'autres discours spécialisés (médecine, droit, finance, etc.). Or comme tout discours, les sciences du langage ont des emplois prédicatifs, argumentaux et actualisateurs dont la description exhaustive et systématique donnera lieu à trois dictionnaires électroniques que l'on peut implémenter dans des logiciels de traitement automatique. La base *Métalangue*, à travers ces deux champs syntactico-sémantiques, est au cœur de cette entreprise descriptive.

S'agissant du champ « concept », il s'agit d'indiquer si le terme est :

- (i) forgé par l'auteur : *sens prédictible* chez D. Corbin, *synapsie* chez Benveniste, etc. ;
- (ii) banalisé au sein de la métalangue des sciences du langage : *mot*, *phrase*, *énoncé* dans des textes qui n'en traitent pas spécifiquement ;
- (iii) repris à un auteur spécifique : *propriété extrinsèque* repris par Cadiot à Anscombe, *dénomination* repris par D. Corbin à G. Kleiber ;
- (iv) ou repris à la métalangue d'un champ disciplinaire : *suffixe* chez D. Corbin, *lexème* chez M.-Fr. Mortureux.

Ces différentes indications sont le plus souvent corrélées au contexte dans lequel apparaît le terme à traiter. Plusieurs marqueurs syntaxiques permettent, en effet, de conclure sur le statut du terme. Par exemple, un marqueur du type *ce qu'on appelle* introduit un terme repris, tandis que des marqueurs du type *on appellera*, *sera appelé* introduisent des termes forgés.

2.2.2 Champs « paratexte »

L'indication des termes apparentés, des termes opposés et du contexte définitoire constitue les deux champs relatifs au paratexte. Aussi bien les termes opposés que les termes apparentés sont distingués selon qu'ils se rapportent au niveau intratextuel ou intertextuel. Cela permet de distinguer les oppositions et les parentés à l'intérieur d'un même texte, et celles qui sont repérables entre les textes. Chaque entrée fournit donc un réseau de relations terminologiques qui dépend en grande

4 Plusieurs termes connaissent cette polysémie systématique : *métonymie*, *métaphore*, *inférence*, etc. La morphologie prend parfois le relais pour exprimer cette dualité entre processus ou propriété et résultat (*méronymie* vs *méronyme* ; *hyperonymie* vs *hyperonyme* ; *néologisme* vs *néologie*, etc.).

partie du contexte d'apparition de l'entrée. Si un terme a deux contextes différents chez un même auteur, il aura par conséquent deux termes apparentés spécifiques à chaque contexte.

Le champ contexte, à travers un extrait de citation où le terme de l'entrée apparaît, indique le voisinage et l'actualisation de l'entrée. Ce champ est d'une importance capitale non seulement par rapport au paratexte, mais par rapport à l'ensemble des champs de la base. On vient de voir que le contexte décide du statut du concept véhiculé par le terme. Rappelons que c'est aussi le contexte qui détermine la fonction primaire des termes (prédicat, argument ou actualisateur). La base *Métalangue* est en réalité un ensemble de termes, tels qu'ils apparaissent dans des contextes définitoires; ce qui pose de fait le problème de pluralité de contextes donnant lieu parfois à des oppositions et des ambiguïtés⁵.

2.2.3 Champs contrastifs

Le champ traduction a un statut particulier. Il s'agit plutôt d'une clé qui permet de diriger l'utilisateur vers la base de la langue mise en correspondance, ici l'arabe. Il ne s'agit donc pas simplement de fournir un équivalent en arabe du terme français, mais d'en donner une description complète comme pour le terme français, avec les mêmes champs, auxquels on ajoutera des champs spécifiques à l'arabe (schème). On retrouve là le principe des dictionnaires monolingues coordonnés qui consiste à superposer deux descriptions complètes, afin d'envisager un transfert automatique de la description en L1 vers celle du terme équivalent en L2 (Buvet 2001).

Ce mode de structuration s'appuie sur une spécificité qui tient au corpus d'étude. L'avantage de la base *Métalangue* est de s'appuyer sur des textes linguistiques traduits. Ont été donc posés en amont des problèmes de conceptualisation, de dénomination, d'idiomaticité, etc. Cette base donne lieu à une réflexion théorique en linguistique contrastive sur des problématiques telles que le poids de la tradition grammaticale (Mejri et Neveu 2004), la cohérence encyclopédique et terminologique à l'intérieur du système et entre deux systèmes L1 et L2.

3 Les corpus et les principes d'extraction

La base *Métalangue* est dans sa seconde année d'existence. Pour cette raison, le corpus reste modeste mais orienté :

- Gaston Gross, *Les expressions figées en français*, Ophrys, 1996;
- Robert Martin, *Comprendre la linguistique*, PUF, 2004;
- Robert Martin, *Pour une logique du sens*, PUF, 1992 (2^e éd.);
- Franck Neveu, *Dictionnaire des sciences du langage*, 2004, Armand Colin;

5 Cela peut être illustré typiquement par l'exemple suivant: « une phrase est un énoncé... ».

- Marie-Françoise Mortureux, *La lexicologie, entre langue et discours*, Colin, 2008 (2^e éd.);
- Alise Lehmann et Françoise Martin-Berthet, *Introduction à la lexicologie*, Colin, 2008 (3^e éd.);
- Denis Apotheloz, *La construction du lexique français*, Ophrys, 2003;
- Maria Teresa Cabré, *La terminologie, Théories, méthodes et applications*, Colin, 1992;
- John Lyons, *Linguistique générale*, Larousse, 1970.

Ont été également retenus des textes initialement publiés en français et traduits en arabe récemment :

- Salah Mejri, Traduction du *Dictionnaire des sciences du langage* de F. Neveu;
- Salah Mejri et Béchir Ouerhani, Traduction de *Les expressions figées en français* de Gaston Gross, 2008;
- Salah Mejri et Taïeb Baccouche, Traduction de *Pour une logique du sens*, 2006.

L'absence de tel ouvrage ou auteur (Le *CLG* de Saussure, Meillet, Beauzée...) ne doit pas être interprétée comme un signe de rejet. Nombre de textes sont numérisés et attendent un traitement. L'orientation vers la formation, la traduction et l'équipement métalinguistique a incité à privilégier des ouvrages dont le coefficient de pénétration dans le milieu universitaire et la recherche est important. Tous les ouvrages entrant dans cet axe ne sont pas encore traités. La représentativité du corpus sera plus grande au fur et à mesure qu'il gagnera en extension, l'attention portant prioritairement sur des textes occupant une position fondamentale dans la bibliothèque linguistique⁶.

Plusieurs principes de sélection des entrées ont été retenus :

- les termes inscrits à l'index des ouvrages, dans les titres de chapitres et d'alinéas. Les unités présentes dans ces paratextes ne se rencontrent pas nécessairement dans le corps du texte (p. ex. *analytique* dans l'index de Martin (2004));
- les séquences marquées typographiquement (gras, italique);
- celles apparaissant dans des distributions métalinguistiques identifiantes (« nous appellerons », « s'appelle », « est le nom de », « signifie », « dénote »...);
- celles constituant elles-mêmes des segments répétés dans un même texte ou disposant d'un constituant récurrent (« suite figée », « séquence figée »).

Une attention particulière est apportée aux variables dans la mesure où il est nécessaire de les baliser comme telles (<statut>/<variante>) afin de signaler l'extension de la reformulation dans un texte. L'identification des variantes s'effectue négativement,

6 Ajoutons que des textes de format moyen ont été privilégiés pour tester les outils d'extraction.

lorsqu'un pilier dénomiatif est signalé comme tel dans le texte (*cf.* plus haut), ou bien lorsque les connaissances du rédacteur de la fiche permettent efficacement la discrimination⁷. On rencontre toutefois des cas où seules des variantes sont coprésentes, p. ex. « expression figurée » et « locution » dans Mortureux (2008).

Certains termes sont rattachés à d'autres par l'index : « analytique » (*cf.* plus haut) renvoie à « démarche analytique » (uniquement dans le titre d'un alinéa), « énoncé analytique », « philosophie analytique », « tendance analytique ». D'autres sont présents dans les notes mais pas dans le texte.

Des outils d'extraction de termes dans les textes sources ont été mis au point au LDI. Ils procèdent par étiquetage et lemmatisation. Ces outils sont paramétrables syntaxiquement. Toutefois, le bruit reste relativement élevé à l'échelle d'un ouvrage entier, ainsi que le temps de post-traitement nécessaire à l'apurement des données. Le reversement des termes déjà collectés dans un dictionnaire de référence tend à réduire par exponentielle le coût d'apurement, sans toutefois en impacter de manière significative la charge. La cause réside pour partie dans le fait que la base ne se réduit pas à de simples listes d'items, des contraintes de retour au texte source étant imposées en vue de la sélection de contextes pertinents. Un concordancier adossé à un lemmatiseur-étiqueteur-extracteur ne suffit pas et ne remplacera pas l'intervention humaine au coup par coup, le contexte décidant dans une majorité de cas de l'intégration ou non d'une séquence. Actuellement un test, prometteur, est effectué à l'aide d'un outil procédant par extraction sémantique. Le bruit est significativement réduit, même si l'inventaire des séquences reste plus centralisé sur des relations explicites. La nomenclature gagne d'autant en densité.

Conclusion et perspectives

La base *Métalangue* dont on vient de présenter le contexte scientifique, les enjeux théoriques et le mode de structuration est doublement orientée vers le passé et vers l'avenir. Partant du constat que la terminologie des sciences du langage souffre d'une instabilité conceptuelle se traduisant par un flou terminologique et une variation sémantique des mêmes termes, dus à l'appartenance à une « jeune » discipline éclatée et hétérogène, traversée par des champs multiples, on a défendu l'idée d'un observatoire et d'un outil de gestion de ce déficit et de cette instabilité, dédié aux chercheurs et aux étudiants. Cet outil n'entend pas se limiter à observer et à gérer au mieux les problèmes de terminologie accumulés par le passé. On souhaite également ouvrir la réflexion vers de nouvelles problématiques telles que le contrastif et le traitement automatique des langues (TAL).

La base *Métalangue* est donc doublement indispensable à la fois par les solutions qu'elle contribue à mettre en place, et les questions qu'elle permet de soulever.

7 P. ex. « mot nouveau » ou « création lexicale » relativement à « néologisme ».

Bibliographie

- Buvet (P-A.), 2001 : « Représentations métalinguistiques de phrases à partir de transducteurs », dans *Revue Informatique et Statistique dans les Sciences Humaines* 36, p. 86-99.
- Gaudin (F.), 1993 : *Pour une socioterminologie*, Publications de l'Université de Rouen, n° 182.
- Gross (G.), 1996 : *Les expressions figées en français*, Ophrys.
- Humbley (J.), 2001 : « Quelques enjeux de la dénomination en terminologie », dans Bosredon (B.), Petit (G.) et Tamba (I.), éd., *Cahiers de praxématique* n° 36, Université Paul Valéry, Montpellier, p. 117-139.
- Kleiber (G.), 1984 : « Dénomination et relations dénominatives », dans *Langages* n° 76, Paris, Larousse, p. 77-94.
- Mejri (S.) et Neveu (F.), 2007 : *La tradition grammaticale*, dans *Langages*, vol. 167, n° 3.
- Martin (R.), 2004 : *Comprendre la linguistique*, Presses universitaires de France.
- Mortureux (M.-Fr.), 2008 (pour la seconde éd.) : *La lexicologie, entre langue et discours*, Armand Colin.
- Neveu (F.), 2004 : *Dictionnaire des sciences du langage*, Armand Colin.
- Petit (G.), 2000 : « Dénomination et lexique », dans *Journal of French Language Studies*, Cambridge, Cambridge University Press.

Salah Mejri,
LDI – UMR 7187 – CNRS – Paris 13,
TIL (00/UR/0201) – Université de la Manouba,
LIA – CNRS – Langue, Traduction, Apprentissage,
Villetaneuse,
France,
smejri@ldi.univ-paris13.fr.

La traduction d'un dictionnaire des sciences du langage : étude de cas

Résumé

Il s'agit de montrer que la traduction d'un dictionnaire des sciences du langage du français vers l'arabe ne peut se faire sans tenir compte des exigences de la macro-structure générale et de la structuration des articles. Parmi les points retenus, on évoque notamment la posture épistémologique de l'auteur, le traitement des citations et des exemples, la place accordée à la tradition grammaticale, l'état des connaissances et des terminologies dans le domaine dans L_1 (langue source) et L_2 (langue cible), etc.

Mots-clés : traduction ; terminologie ; épistémologie ; dictionnaire ; langue spécialisée ; métalangue.

1 Problématique générale

L'article d'un dictionnaire encyclopédique serait, selon la *doxa*, une sorte d'équivalence qui repose sur une relation univoque entre le terme qui lui sert d'entrée et le contenu conceptuel développé dans le corps de l'article. Ainsi, entrée et développement seraient les deux termes d'une relation qui se présente comme une donnée stable, normée et, par conséquent, suffisamment autonome pour permettre une traduction séparée des articles du dictionnaire.

Cette vision repose sur une illusion d'optique qui ne tient pas compte d'au moins cinq contraintes qui rendent cette autonomie toute relative :

- la macro-structure du dictionnaire qui est censée régir son économie générale ;
- la structuration interne des articles qui tient compte de la taille du développement des contenus et de l'ensemble des éléments structurant ce développement ;
- la posture épistémologique de l'auteur du dictionnaire, c'est-à-dire les choix théoriques implicites ou explicites sous-jacents à la conception générale du

dictionnaire (choix des entrées, des auteurs, des citations, etc.; analyses; traitement de la tradition grammaticale, etc.);

- l'état des connaissances dans la discipline en L₂, de la terminologie et des descriptions disponibles.

Nous essayerons de développer dans ce qui suit ces cinq points en puisant les exemples dans la traduction en arabe du *Dictionnaire des sciences du langage* de F. Neveu.

2 La macro-structure du dictionnaire

Un dictionnaire spécialisé est un ouvrage qui est censé refléter l'état des connaissances dans un domaine précis à un moment de son évolution. Le traduire ne peut se faire indépendamment de son architecture générale, dont nous retenons particulièrement:

- a) La taille de la nomenclature et celle du dictionnaire: plus la taille est grande, plus le contenu est analytique, c'est-à-dire plus il est distribué entre les entrées. Plus la taille est petite, plus le contenu est synthétique. La traduction n'est plus la même selon que la matière est condensée ou diluée;
- b) Le classement des entrées par champ disciplinaire: que ce classement soit explicite ou implicite, il conditionne la structuration de la matière lexicographique, et ce, au moyen du marquage des domaines, des indications relatives aux options théoriques et au positionnement par rapport au cœur de la spécialité. Le marquage du domaine inscrit l'article du dictionnaire dans un réseau conceptuel précis, celui de la phonétique, de la syntaxe, de la sémantique, etc. Il s'ensuit que les articles de chaque discipline se trouvent dans une relation d'interdépendance. La mention des écoles ou des théories, comme la psychomécanique ou la grammaire générative, implique nécessairement des choix théoriques dont il faut tenir compte lors de la traduction. La position que chaque article a par rapport au cœur de la spécialité impose des précautions de nature terminologique et conceptuelle, notamment pour tous les articles qui sont à la périphérie de la discipline, et par conséquent en rapport avec d'autres disciplines ayant leurs propres terminologies et leurs appareils conceptuels, comme c'est le cas par exemple pour l'épistémologie, la neurologie et les pathologies du langage;
- c) Le système des renvois: il ne s'agit pas d'une simple technique éditoriale. De tels renvois favorisent une lecture en hypertexte et permettent de dégager une hiérarchie entre les articles. Si on prend comme exemple l'entrée *adjectif*, on est renvoyé à l'article *partie du discours*. Ce renvoi implique au moins deux choses: la présence du contenu conceptuel de l'adjectif dans l'article réservé aux parties du discours et le réseau de relations dans lequel s'insère l'entrée *adjectif* (nom, adverbe, verbe, etc.). Les deux articles *Principes et paramètres* et *bit* illustrent les points que nous venons de mentionner (cf. annexe 1).

3 La micro-structure du dictionnaire

Le développement interne de chaque article obéit nécessairement à l'architecture générale du dictionnaire, et ce, à travers plusieurs éléments qui structurent l'élaboration de l'article :

- a) l'indication du domaine assure un lien avec tout l'appareil conceptuel de ce domaine, ce qui détermine en grande partie la valeur de chaque terme et dissocie les différents emplois du même terme :

إسمانيّة Nominalisme [فلسفة اللغة، المنطق، الدلاليّة]
 NOMINALISME [philosophie du langage, logique, sémantique]

- b) les renvois aux autres articles du dictionnaire créent un réseau conceptuel sous lequel le contenu de chaque article ne peut être saisi indépendamment de celui des termes qui font partie de la structure du concept décrit dans l'article. On peut ramener les renvois à trois grands ensembles :
 - b.1. un ensemble qui comporte les termes définitoires du concept décrit dans l'article, qu'ils soient hyper- ou hyponymes. C'est le cas de l'article *inférence* qui renvoie à des concepts définitoires de l'inférence comme l'implication, l'implicature, l'induction, etc. (cf. annexe 2) ;
 - b.2. un ensemble qui comporte les termes constitutifs d'un champ disciplinaire dont les contenus se construisent les uns par rapport aux autres. L'article consacré aux pathologies du langage offre l'occasion de voir comment un article peut comporter presque la totalité des renvois aux entrées couvrant tout le champ (cf. annexe 3) ;
 - b.3. un ensemble qui sélectionne les réseaux terminologiques propres à des théories. L'exemple de l'article sur l'incidence, terme propre à la psychomécanique, montre qu'à travers les renvois on peut récupérer des termes spécifiques à la même théorie (cf. annexe 4) ;
- c) l'appartenance du même terme à deux ou davantage de domaines, ce qui établit des liens d'un type particulier entre les articles qui comportent des renvois à plusieurs paradigmes. Nous renvoyons par exemple à l'article *période* qui spécifie qu'un tel terme s'emploie en rhétorique et en macrosyntaxe, à celui de *clause* qui indique que les domaines où ce terme s'inscrit sont la narratologie et la macrosyntaxe ;
- d) les citations des auteurs : il s'agit le plus souvent de citations définitoires (cf. par exemple l'article *exophrastique*), illustratives (cf. l'article *calcul*), ou contradictoires (cf. l'article *idiolecte*) (cf. annexe 5). Leur traduction exige de prendre au moins trois précautions :
 - d.1. celle qui tient compte des relations entre la citation et le corps de l'article dans lequel elle est insérée ;

- d.2. celle qui est relative aux liens qu'elle entretient avec les citations du même auteur dans le dictionnaire et avec celles des autres auteurs;
- d.3. celle qui prend en considération les rapports qui existent entre la citation et le texte dont elle est extraite;
- e) les exemples jouent un rôle central dans les textes linguistiques parce qu'ils servent de support aux commentaires métalinguistiques: leur présence justifie les commentaires et vice-versa. D'où les trois possibilités de traductions suivantes:
 - e.1. soit on conserve les exemples d'origine et les commentaires qui y réfèrent, comme c'est souvent le cas pour les exemples concernant les langues rares et les cas idiosyncrasiques;
 - e.2. soit on remplace les exemples portant sur la langue source par des exemples de la langue cible et on développe les commentaires adéquats aux nouveaux exemples. Dans ce cas, on fait figurer en note des exemples d'origine ou l'on opte pour la présentation inverse;
 - e.3. soit on substitue aux exemples du texte de L1 des exemples de L2 tout en conservant les mêmes commentaires: ce cas concerne les faits de linguistique générale ou les faits partagés par L_1 et L_2 ;
- f) les indications étymologiques de l'auteur du dictionnaire: il s'agit nécessairement d'indications propres à la langue source. Leur traduction peut paraître inutile. Mais si on les traduit, on fournit au lecteur des éléments relatifs à la construction morphologique de la terminologie de L_1 ; ce qui peut représenter l'une des sources possibles de création terminologique en L_2 .

4 La posture épistémologique de l'auteur du dictionnaire

C'est le point le plus difficile à saisir. Si on part du principe que le traducteur doit respecter l'intégralité du contenu du texte de L_1 , il doit saisir la posture épistémologique de l'auteur et faire en sorte qu'elle soit présente dans le texte d'arrivée. On peut ramener une telle posture aux points suivants:

- a) le point de vue adopté par rapport à l'objet décrit: cela concerne la manière dont l'auteur conçoit son entreprise lexicographique. Pour F. Neveu par exemple, il ne s'agit pas de faire un dictionnaire qui obéit à ce qu'il appelle « les trois vertus déclarées d'une bonne terminologie »: l'économie, la transparence et la cohérence. Son objectif est plutôt de « refléter un état de la discipline [...] observable à partir de son vocabulaire, c'est-à-dire partir des pratiques terminologiques effectives »;
- b) le décroisement de la discipline: souvent les dictionnaires spécialisés focalisent sur ce qui est considéré comme le cœur de la discipline. Le dictionnaire de F. Neveu s'inscrit dans le choix inverse. C'est pourquoi il a introduit plusieurs entrées et concepts appartenant à des disciplines connexes comme l'informatique, l'histoire, l'épistémologie, etc.;

- c) la part des théories et des auteurs dans la construction du dictionnaire: la posture épistémologique de l'auteur se vérifie également à travers les choix des écoles, tendances et théories (psychomécanique, pragmatique, grammaire générative, etc.) et celui des auteurs les plus cités (G. Guillaume, R. Martin, S. Auroux);
- d) le traitement réservé à la tradition grammaticale: le rapport à la tradition grammaticale s'avère l'un des plus difficiles à saisir. Cela s'explique par la complexité des relations qu'entretiennent les sciences du langage avec l'héritage conceptuel et terminologique. L'auteur du dictionnaire est censé tenir compte de cette dimension historique, et ce, en établissant un parallèle entre les concepts de la tradition et les usages qu'on en fait actuellement, et en précisant le flou conceptuel et méthodologique qui accompagne le recours ambivalent à la tradition qui fait qu'on s'en sert comme argument d'autorité et qu'on la critique incessamment parce qu'on la considère comme un vecteur de conservatisme et de désuétude;
- e) la présentation de l'état des sciences du langage dans sa diversité, ses contradictions et ses incohérences: la tâche du traducteur est très délicate de ce point de vue. Cela exige qu'il fasse en sorte que le lecteur de L_2 ne voie pas dans ce choix une source de contradiction et d'incohérence de l'auteur lui-même.

5 Le degré de stabilité conceptuelle et terminologique dans la discipline en L_1 et L_2

À ce propos, deux constats s'imposent: il est rare d'avoir le même degré de stabilité conceptuelle et terminologique dans deux ou plusieurs langues; toute construction conceptuelle dans un domaine scientifique porte les empreintes de la langue dans laquelle cette construction a eu lieu. Partant de ces constats, le traducteur doit tenir compte de:

- a) de l'asymétrie (ou de la dissymétrie) terminologique entre L_1 et L_2 . Dans le cas du français et de l'arabe, cela se constate au niveau de la masse terminologique générale. Le français dépasse en cela l'arabe; ce qui a comme résultat direct la création terminologique et tout ce qui s'ensuit comme problèmes de normalisation. Cela se vérifie par exemple pour la terminologie du TAL, de certaines théories ou de certains domaines. Cela se vérifie également au niveau de la part terminologique qui revient à la tradition grammaticale dans l'usage actuel: dans ce domaine l'arabe l'emporte sur le français, notamment en syntaxe, rhétorique et morphologie;
- b) de l'écart au niveau de la conceptualisation: un concept n'a d'existence qu'à travers le terme qui lui sert de signification et tous les discours élaborés pour en décrire la structuration, le contenu et les fondements épistémologiques. Pour qu'il y ait transfert de concepts, si de tels discours n'existent pas en L_2 , il faut éviter le risque de fournir une représentation du concept tronquée de son ancrage dans la discipline et dans la langue dans laquelle il a été conçu.

Pour la traduction du *Dictionnaire des sciences du langage*, nous avons tenté de réduire la dissymétrie à travers les traductions réalisées par notre équipe de plusieurs ouvrages traitant de la syntaxe, de la sémantique, de la logique, de la lexicologie, etc. Mais d'autres domaines ont été fortement marqués par l'absence de discours disponibles en arabe comme c'est le cas pour le traitement automatique des langues.

Conclusion

Pour que la traduction de ce type de dictionnaire ne tombe pas dans la littéralité lors du passage du L_1 à L_2 , il faut :

- situer les articles, pris séparément, dans une perspective structurelle qui tient compte des choix macro- et micro-structurels ; leur traduction ne peut ignorer l'ensemble des liens qui rattachent chaque article au reste du dictionnaire ;
- voir dans ces liens un ensemble de réseaux qui sont de nature à la fois terminologique, conceptuelle et théorique ;
- essayer de rendre compte du mieux qu'on peut de la posture épistémologique de l'auteur.

Après de telles précautions, peut-on être sûr que le transfert du contenu du dictionnaire pourrait être total ? La réponse est évidemment négative. Toute traduction est nécessairement une question de déficit, déficit qui ne s'explique pas exclusivement par la qualité de la traduction, mais, comme on a essayé de le signaler, qui trouve son origine dans l'état général de la discipline en L_1 et L_2 . Pour que l'évaluation de la traduction soit objective, il ne faut pas se limiter, comme on l'a fait ici, aux choix de l'auteur du texte source, il faut étudier également les choix du traducteur ; ce qui fera l'objet d'un autre texte.

Bibliographie

Ouvrages traduits

- Mejri (S.) et Baccouche (T.), 2006 : Traduction en arabe de l'ouvrage de Martin (R.), 1992 : *Pour une logique du sens*, PUF, 2^e éd., Publications de l'AOT, Beyrouth.
- Mejri (S.) et Ouerhani (B.), 2008 : Traduction en arabe de l'ouvrage de Gross (G.), 1996 : *Les expressions figées en français*, Ophrys, Publications du CERES.
- Mejri (S.), 2010 : Traduction en arabe de Neveu (F.), 2004 : *Dictionnaire des sciences du langage*, Armand Colin, 2004, Publications de l'AOT, Beyrouth.

Articles

- Balliu (Chr.), 2002 : *Les traducteurs transparents. La traduction en France à l'époque classique*. Bruxelles, Les Éditions du Hazard, Préface d'André Clas (Université de Montréal).
- Buvet (P.-A.) et Blanco (X.), 1999 : « À propos de la traduction automatique des déterminants de l'espagnol et du français », dans *Meta* 44 (4), p. 525-545.

- Goffin (R.), à paraître: « Interlinguistique et Terminologie. Un bref regard en arrière », dans *Meta*, 55-1, Hommages au Professeur André Clas, Presses de l'Université de Montréal.
- Ladmiral (J.-R.), 1994 et 2002, *Traduire: théorèmes pour la traduction*, Paris, Payot, 1979 (Petite Bibliothèque Payot, n° 366). Rééditions (avec une pagination identique), augmentées d'une préface (p. V- XXI): Paris, Gallimard, (coll. « Tel », n° 246).
- Ladmiral (J.-R.), à paraître: « Sur le discours méta-transductif de la traductologie », dans *Meta*, 55-1, Hommages au Professeur André Clas, Presses de l'Université de Montréal.
- Mejri (S.) et Baccouche (T.), (dir.), 2000 : « La traduction dans le monde arabe », dans *Meta*, numéro spécial 45 (3), Presses universitaires de Montréal.
- Mejri (S.) et Mogorron Huerta (P.), (dir.), 2009 : *Figement, défigement et traduction*, Université d'Alicante.
- Mejri (S.), 2008a: « La traduction des textes spécialisés: le cas des sciences du langage », dans *Traduire: un métier d'avenir*, Balliu (Chr.), dir., vol 2, Les éditions du Hazard, Bruxelles, p. 119-144.
- Mejri (S.), 2008b, (dir.): *Equivalence* n° 35, *La traduction de l'humour et des jeux de mots*, Éditions du Hazard, Bruxelles.
- Mejri (S.), 2008c, (dir.): *Meta* 53 (2), *La traduction des séquences figées*, Les Presses de l'Université de Montréal.
- Mejri (S.), 2010, « Les dénominations syntagmatiques dans les SHS: collocations ou séquences figées », dans *La phraséologie dans les discours spécialisés*, Mejri (S.) et Gautier (L.), dir., Presses Universitaires de Bourguignon (sous presse).

Annexes

Annexe 1

| | |
|--|--|
| <p>PRINCIPE S ET PARAMÈTRE S [syntaxe]</p> <p>Cadre théorique d'une des dernières étapes de la grammaire générative. La notion de principes vise à formuler l'hypothèse d'une grammaire universelle qui conditionnerait l'acquisition des langues (inscrite dans le dispositif biologique de l'espèce humaine), et qui définirait des règles de fonctionnement linguistique partagées par toutes les langues. La notion de paramètres sert à déterminer les espaces de variation possible des langues particulières relativement à ces principes, c'est-à-dire leur mode d'instanciation spécifique de la grammaire universelle. Ainsi le principe de projection qui s'inscrit dans la théorie X-barre, fait apparaître des invariants (tout syntagme est décrit comme la projection maximale d'une tête qui sélectionne un spécifieur* et un complément*). À ce principe correspondent des paramètres variables selon les langues. Par exemple, en français l'application de ce principe fait apparaître l'ordre spécifieur + tête et tête + complément. Le formalisme de la théorie X-barre permet de représenter, pour une structure déterminée, ce qui relève des principes et ce qui relève des paramètres.</p> <p>► Voir <i>Liage, Transformation</i>.</p> <p>► Chomsky N., 1982, <i>Some Concepts and Consequences of a Theory of Government and Binding</i>, MIT Press ; trad. franç. <i>La Nouvelle Syntaxe</i>, Éditions du Seuil, 1987. ► Chomsky N., 1991, <i>Théorie du gouvernement et du liage : les conférences de Pise</i>, Éditions du Seuil.</p> | <p>مبادئ ومعايير [التركيبية] Principles et paramètres</p> <p>هي الإطار النظري لإحدى آخر مراحل النحو التوليدي. يهدف مفهوم المبادئ إلى التعبير عن فرضية النحو الكوني الذي من المحتمل أنه يتحكم في اكتساب الألسن (المسجل في الجهاز البيولوجي للجنس البشري)، وأنه قد بعرف قواعد الاستعمال اللساني المشترك بين جميع الألسن. ويُستعمل مفهوم المعايير لتحديد فضاءات الإبدال الممكن للألسن الخاصة بالنسبة إلى هذه المبادئ، أي صبغة تحققها الخاص للنحو الكوني. هكذا يُبرز مبدأ الإسقاط المسجل في نظرية س-بار توابت (كل مركب يقع وصفه باعتبار إسقاط أقصى لرأس بنقي مخصصاً ومتمماً*). وتقال هذا المبدأ توابت معتبرة حسب الألسن. مثال ذلك أن تطبيق هذا المبدأ في الفرنسية يُبرز التركيب مخصص+رأس ورأس+ متمم. تسمح شكلية نظرية س-بار، بالنسبة إلى بنية مختلفة، بتمثل ما يتعلق بالمبادئ وما يتعلق بالتوابت.</p> <p>◀ انظر : ربط تحويل .</p> <p>◀ المراجع :</p> <p>► Chomsky N., 1982, <i>Some Concepts and Consequences of a Theory of Government and Binding</i>, MIT Press ; trad. franç. <i>La Nouvelle Syntaxe</i>, Éditions du Seuil, 1987. ► Chomsky N., 1991, <i>Théorie du gouvernement et du liage : les conférences de Pise</i>, Éditions du Seuil.</p> |
| <p>BIT [informatique]</p> <p>Acronyme formé sur l'anglais <i>binary digit</i>, « élément discret binaire ». Le bit est la plus petite unité d'information stockée par la mémoire d'un ordinateur. Cette unité peut prendre deux valeurs : 0 et 1. La combinaison de plusieurs bits permet de représenter n'importe quel type d'information.</p> <p>► Voir <i>Octet</i>.</p> | <p>بيت [الإعلامية] BIT</p> <p>[الكلمة الفرنسية اقتضاب جزئي من الإنكليزية <i>Binary Digit</i> "عنصر متواصل ثنائي"]. والبيت ه وأصغر وحدة معلوماتية يقع تخزينها في ذاكرة الحاسوب. ويمكن أن تكون لهذه الوحدة قيمتين : 0 و1. ويسمح التوليف بين العديد من وحدات التعداد بتمثيل أي نمط من المعلومات.</p> <p>◀ انظر : هائية</p> |

Annexe 2

| | |
|--|--|
| <p>INFÉRENCE [logique, sémantique, pragmatique]</p> <p>Du latin <i>inferre</i>, « porter dans, mettre en avant, produire ». Une inférence est un processus selon lequel une proposition* est admise en vertu de son lien logique avec une ou plusieurs propositions antécédentes tenues pour vraies (prémises). La notion d'inférence entre dans le champ des sciences du langage en ce qu'elle fournit une consistance logique à l'implicite* du discours. Robert Martin (<i>Inférence, antonymie et paraphrase</i>, 1976) a distingué deux types principaux d'inférences : les inférences nécessaires, indépendantes de la situation d'énonciation, et inscrites dans la structure linguistique de l'énoncé (l'implication*, le présupposé*); les inférences possibles, correspondant à un implicite pragmatique, et déterminées par des paramètres situationnels (le sous-entendu*).</p> <p>➤ Voir <i>Coopération, Déduction, Implication, Implicature, Implicite</i> [A], <i>Induction, Moteur d'inférence, Présupposé, Proposition, Sous-entendu</i>.</p> | <p>استدلال [المنطق، القلائبة، القائلية] <i>Inférence</i></p> <p>[أصل الكلمة الفرنسية من اللاتينية، <i>inferre</i>، "حمل في، قدم، أنتج".] للاستدلال عملية بموجبها يُقبل قول* بموجب علاقته المنطقية مع قول أو مجموعة أقوال سابقة تُعتبر حقًا (المقدمات). يدخل مفهوم الاستدلال مجال علوم اللغة من حيث أنه يوفّر حلقة منطقية لضميمة* الخطاب. ويميّز رولان مارتان (الاستدلال والنضاد والصنوع، 1976) بين نوعين رئيسيين من الاستدلالات: الاستدلالات الضرورية، وهي مستقلة عن وضعية اللفظ ومستقلة في البنية السلسلية لللفظ (لافتضاء*، الافتراض المسبق*، والاستدلالات الممكنة التي توافق ضميمة شاولية والتي تُحدّد بواسطة مغليس وضعية (المقدّر*).</p> <p>◀ نظر: استنتاج، استهزاء، إهانة، إهتزاز، افتخارية، افتراض مسبق [A]، تعاون، ضميمة [A]، قانون الخطاب، قول، معدرك استدلال، مهذّر.</p> |
|--|--|

Annexe 3

| | |
|---|--|
| <p>PATHOLOGIE DU LANGAGE [neurolinguistique]</p> <p>On appelle <i>pathologie du langage</i> un ensemble de dysfonctionnements neurophysiologiques de l'activité langagière susceptibles d'affecter tout à la fois l'appareil phonatoire et auditif. On distingue notamment entre les troubles du langage, ou aphasies*, intervenant à la suite d'une lésion locale du système nerveux central, et les troubles de la parole (comme le bégaiement), qui se manifestent par un dysfonctionnement spécifiquement moteur ou sensoriel des organes périphériques de la parole.</p> <p>➤ Voir <i>Agrammatisme, Anarthrie, Aphasie, Aphémie, Associationnisme, Agraphie, Alexie, Dysarthrie, Dysgraphie, Dysphasie, Dysprosodie, Écholalie, Logatome, Neurolinguistique, Paragrammatisme, Paragraphie, Paraphasie</i>.</p> | <p>مرض اللغة Pathologie du langage [التسليكات العصبية]</p> <p>يطلق مرض اللغة على مجموعة من الاضطرابات المختلة العصبية-الفيزيولوجية في النشاط اللغوي، ومن شأنها أن تسبّب في ذات الوقت الجهل التصويبي والسحبي. ويقع التمييز بالخصوص بين اضطرابات اللغة أو الحصاة* التي تحصل على إثر ضرر محلي في النظام العصبي المركزي، واضطرابات الحديث (مثل اللجلجة) التي تظهر من خلال خلل في الاشتغال الحركي أو الحسي للأجهزة المحيطة للحديث بالخصوص.</p> <p>◀ انظر: حبة نحوية، حبة نطوية، حبة، حبة كلامية، انصافية، حبة كتابية، حبة قرائية، لغة، اضطراب كتابي، اضطراب خطابي، اضطراب نغمي، تردد، متقطع، لسانيات عصبية، اضطراب نحوي، عيب خطي، عيب استبدالي.</p> |
|---|--|

Annexe 4

INCIDENCE [syntaxe, sémantique, psychomécanique]

À partir du latin scolastique *incidere*, « tomber sur, survenir », du latin classique *cadere*, « tomber ». La notion d'incidence a été principalement développée par Gustave Guillaume, dans le cadre de la théorie de la psychomécanique* du langage, où elle est employée dans deux domaines distincts.

Dans l'étude du temps* verbal, l'incidence, en opposition à la décadence, définit un des moments successifs qui caractérisent le mouvement d'échéance de l'événement au temps d'univers : celui du procès en accomplissement où l'événement est vu *arrivant* au temps. La décadence définit, inversement, le moment du procès accompli où l'événement est vu *arrivé* au temps. La langue française rend compte de cette distinction dès le mode nominal par l'opposition qu'elle fait percevoir entre l'infinitif (ex. *marcher* : niveau d'incidence), le participe présent (ex. *marchant* : niveaux d'incidence et de décadence), et le participe passé (ex. *marché* : niveau de décadence). Mais les niveaux d'incidence et de décadence se retrouvent à chaque étape de la chronogénèse*.

La notion d'incidence est par ailleurs développée dans le cadre de l'étude des mécanismes syntaxiques qui régissent le fonctionnement dans la phrase des parties* du discours. Pour Guillaume le mouvement d'incidence :

[...] a trait au mouvement, absolument général dans le langage, selon lequel, partout et toujours, il y a apport de signification et référence de l'apport à un support.

G. Guillaume, *Leçons de linguistique, 1948-1949, Série B, Psycho-systématique du langage—Principes, méthodes et applications I*, Klincksieck, 1971.

انعكاس] Incidence التركيبية، الدلالية، النفسية الآتية]

أصل الكلمة الفرنسية من اللاتينية المدرسية، من *incidere* "سقط على، حدث"، من اللاتينية الكلاسيكية *cadere* "سقط". لقد طوّر مفهوم الانعكاس أساساً فوستاف فيوم في إطار نظرية نفس-آتية* اللغة حيث يستعمله في مجالين متميزين.

في دراسة الزمن الفعلي* يعرف الانعكاس، في تقابل مع الانحلال، أخذ الأوقات المتعاقبة التي تميز حركة تحول الحدث في زمن الكون: حركة العمل أثناء الإنجاز حيث يرى الحدث أثناء حلوله في الزمن. وبالعكس يعرف الانحلال وقت العمل المنجز حيث يرى الحدث وقد حلّ في الزمن. يقوم اللسان الفرنسي بهذا التمييز انطلاقاً من الصيغة الاسميه وذلك بالمقابل التي تلاحظها بين الفعل المتصرف (مثال:

marcher [يمشي]: مستوى الانعكاس) ومشارك الحاضر (مثال: *marchant* [هو يمشي]: مستوى الانعكاس والانحلال) ومشارك الماضي (مثال *marché* [قد مشى]: مستوى الانحلال). إلا أن مستوى الانعكاس والانحلال يوجدان في كل مرحلة من مراحل التوليد الزمني*.

وفضلاً على ذلك وقع تطوير مفهوم الانعكاس في إطار دراسة الأليات التركيبية التي تتحكم في اشتغال أقسام الكلام في الجملة. فبالنسبة إلى فيوم حركة الانعكاس:

[...] لها صلة بالحركة العامة إطلاقاً في اللغة، التي بموجبها يكون هناك دائماً وحيث ما كان منقول من الدلالة وإحالة للمنقول على ناقل.

ق. فيوم، *دروس في اللسانيات 1948-1949*، سلسلة ب، نفس-نظامية اللغة-مبادئ ومناهج وتطبيقات I، Klincksieck، 1971.

Annexe 5

| | |
|--|--|
| <p>EXOPHRASTIQUE [syntaxe, sémantique]</p> <p>À partir du grec <i>εξω</i>, « au-dehors ». Le terme a été introduit dans la description linguistique par Claude Guimier (<i>Syntaxe de l'adverbe anglais</i>, 1988 ; <i>Les Adverbes du français – Le cas des adverbes en -ment</i>, 1996) pour caractériser le fonctionnement de certains adverbes* usuellement décrits comme adverbes de phrase dans la grammaire traditionnelle.</p> <p>[...] l'adverbe est dit exophrastique lorsqu'il est le signe d'une idée regardante portée par l'énonciateur sur tout ou partie de son énoncé. À ce titre, l'adverbe exophrastique, extra-prédicatif par nature, est un élément extérieur à l'énoncé et ne participe pas à la construction du contenu propositionnel. Il joue lui-même le rôle d'une proposition, réduite à un mot (d'où l'appellation d'adverbe-phrase qui a été proposée), qui commente de l'extérieur une phrase ou un syntagme. L'adverbe exophrastique peut en effet être incident à une phrase complète (d'où l'étiquette « adverbe de phrase » communément utilisée par les grammairiens) : Apparemment, il est venu sans rien dire à personne (l'adverbe est incident à toute la phrase moins l'adverbe lui-même), comme il peut être incident à une portion de phrase, à un syntagme : Il est venu sans rien dire à personne apparemment (l'adverbe est incident au groupe prépositionnel sans rien dire à personne).</p> <p>C. Guimier, <i>Les Adverbes du français – Le cas des adverbes en -ment</i>, Ophrys, 1996.</p> <p>► Voir Adverbe, Endophrastique.</p> | <p>خارج - جملي Exophrastique [التركيبية، الالائية]</p> <p>[أصل الكلمة الفرنسية يوناني، من <i>εξω</i> "خارج"] كلود غيمبي (تركيبية الوديف الإنكليزي، 1988 ؛ الودائف في الفرنسية. حالة الودائف المنسقة بـ <i>-ment</i>، 1996) هو الذي أدخل هذا المصطلح في الوصف اللساني لوسم استغلال بعض الودائف* التي توصف عادة في النحو التقليدي لودائف جمل.</p> <p>[...] يُعتبر الوديف خارج جملي إذا ما كان علامة الفكرة الخارجة التي يحملها المشتغل على فوطه جملة أو تعديلاً. وهكذا فإن الوديف الخارج جملي، وهو بطبيعته خارج - إنشائي يشكل عنصراً خارجاً عن الفط لا يساهم في بناء المحتوى القولي. و هو يندب في حد ذاته دور قول ويق اختزاه في كلمة (من هنا تسمية ريف- جملة المقترحة) يقوم بقول جملة من الخارج أو اختزالها في مركب. وفقاً فإن الوديف الخارج جملي يمكن أن يقع انعكسه على جملة قلّة (من هنا تسمية " ريف الجملة" الشاواول عند الشحلة) : على ما يبدو. لهذا ريف جملي أو يعلم بذلك أحد (انعكس الوديف يقع على كلّ الجملة ما عدا الوديف ذاته) كما أنه يمكن انعكسه على جزء من الجملة أي مركب : لهذا جاء. ريف جملي أو يعلم بذلك أحد، على ما يبدو (الوديف يقع انعكسه على المركب الخرفي ريف أو يعلم بذلك أحد)</p> <p>لر قيمي، الودائف في الفرنسية. حالة الودائف المنسقة بـ <i>-ment</i>، Ophrys، 1996.</p> <p>◀ انظر: داخل جملي، ريف</p> |
|--|--|

CALCUL [sémantique formelle, intelligence artificielle]

Le terme de *calcul* est employé de diverses manières en linguistique, et parfois métaphoriquement pour exprimer le désir d'une épistémologie de la science du langage fondée sur la mathématisation* de son objet.

En dépit des apparences, et des habitudes acquises d'un système éducatif qui corréle systématiquement l'étude des langues à celle des lettres, l'activité calculatoire est une activité naturelle en grammaire, illustrée notamment par l'usage des paradigmes, qui n'est pas sans relation avec un traitement algorithmique du langage.

Dans la science linguistique contemporaine, le calcul a vocation à désigner, en sémantique formelle, où est exploité le métalangage de la logique, les opérations destinées à rendre compte de la vériconditionnalité* du sens. Dans cette perspective, il sert notamment à l'étude de la portée des quantificateurs (*chaque, tous, le, un* etc.), susceptibles de produire une ambiguïté* sémantique. Le calcul des prédicats, qui s'inscrit dans une approche logique de la quantification*, procède à l'analyse des propositions en les considérant comme des propriétés attribuées à des arguments. La forme logique devient alors un module fondamental de l'interprétation du sens.

L'intelligence* artificielle et le traitement* automatique du langage naturel ont tiré parti des travaux sur la calculabilité du mathématicien Alan Turing, qui a proposé dans le courant des années 1930 un ensemble de réflexions sur les nombres calculables, élaborant le concept d'un mécanisme (dit *machine de Turing*) permettant de définir la calculabilité de manière non intuitive. À la même époque se sont développées les recherches sur le lambda-calcul (λ -calcul), mis au point par Alonzo Church, largement exploitées par la sémantique formelle contemporaine :

Le lambda-calcul procède de la notion primitive de fonction : les arguments des fonctions sont aussi des fonctions. Cette primauté a été rendue possible par la réduction des fonctions à plusieurs arguments à des fonctions unaires [...]. Ainsi une fonction f à deux variables, p et q , s'écrira non pas selon la notation de Russell $f(p, q)$, mais $(fp)q$, où l'on envisage que la fonction fp s'applique à q ; $f(p, q, r)$ devient $((fp)q)r$, etc. L'invention la plus importante est l'opérateur d'abstraction, noté avec la lettre grecque λ (lambda). C'est cet opérateur qui permet d'exprimer la substitution. [...] le λ -calcul permet [...] d'obtenir une caractérisation générale de la calculabilité.

S. Auroux & alii, *La Philosophie du langage*, PUF, 1996.

Le terme de *calculabilité* est également employé en pragmatique pour caractériser un type d'implicature* conversationnelle.

► Voir *Algorithme, Computationnisme, Formalisme, Mathématisation, Vériconditionnalité*.

حساب Calcul [الآلاتية الشكلية، الذكاء الاصطناعي]

يُستعمل مصطلح حساب بطرق مختلفة في اللسانيات ويستخدم في بعض الأحيان بصفة استعارية للتعبير عن الرغبة في علمية علم لغة يرتكز على تربيص* موضوعه.

ورغم السطوح والحادات الشكلية في نظام تربيص يرتبط آتيا دراسة الأكنس بدراسة الآداب فإن النشاط الاحصائي في النحو هو نشاط طبيعي يجسد بالخصوص في استعمال الجريدات التي ليست بدون علاقة مع معالجة حوارية لغة.

ويستخدم الحساب في علم اللسانيات المعاصر في الآلاتية الشكلية التي تستغل الورقة المنطقية لتعيين عمليات تهدف إلى وصف حقائق* المعنى. وهو يستخدم من هذا المنطلق بالخصوص لدراسة مدى السمكيات (عَل، جمع، الـ الخ) التي من شأنها أن تحدث التباسا* دلاليًا. يقوم حساب المسانيد، الذي ينخرط في مقارنة منطقية للإكمام*، بتحليل الأقوال باعتبارها خصوصيات مستدة إلى معمولات، ويصبح الشكل المنطقي عندئذ وحدة أساسية في تأويل المعنى.

وقد استفاد الذكاء* الاصطناعي والمعالجة* الآتية للغة الطبيعية من أعمال آلان تورينج حول قابلية الاحصاب للرياضي وهو الذي اقترح أثناء السنوات 1930 مجموعة من الآراء حول الأعداد القابلة للاحصاب ميلورا منصور الآتية (تسمى آلة تورينج) تسمح بتعريف قابلية الاحصاب بطريقة لا حدسية.

وتطورت الأبحاث في الفترة نفسها حول الحساب الآتية (λ -حساب) الذي وضعه أونزو تشورنن والذي وقع استغلاله بوفرة في الآلاتية الشكلية المعاصرة :

يقوم الحساب الآتية على المفهوم البدائي للوظيفة : معمولات الوظائف هي أيضا وظائف. وقد أصبحت هذه الأولوية ممكنة بواسطة اختزال الوظائف ذات السمكيات المتعددة إلى وظائف أحادية [...] هكذا تستعمل وظيفة ذات متغيرين، ج و ق، لا حسب ترميز روسال و(ج، ق)، وإنما (و ج) ق؛ حيث نعتبر أن الوظيفة و ج تنطبق على ق، و(ج، ق)، غ تصبح (و ج) ق(ق) غ، الخ. إن أهم ابتكار هو عامل التجريد الذي يستعمل بواسطة الحرف اليوناني λ (لامبدا) إن هذا العامل هو الذي يسمح بالتعبير عن الترميز [...]. يسمح الحساب الآتية [...] بالحصول على رسم عام لقابلية الحساب.

س. أوروا والآخرين، *لسفة اللغة*، PUF، 1996.

ويستخدم مصطلح قابلية الاحصاب كذلك في التداولية لتمييز شرط من الافتضائية* الحوارية.

◀ **انظر** إحصائية، تربيص، حقائقية، حوارية، شكلية.

IDIOLECTE [analyse du discours, sémantique textuelle]

Mot emprunté à l'anglais *idiolect*, qui a été formé sur le grec *idios*, « propre, spécial », et sur le français *dialecte*. La notion d'idiolecte sert à désigner un usage de la langue marqué par les dispositions particulières de l'énonciateur. Plus précisément, on appelle *idiolecte* les normes ou les régularités individuelles dont témoignent ces dispositions.

Le faible développement de la notion d'idiolecte en sciences du langage et la permanence de la problématique qu'elle soulève révèlent de manière significative la difficulté de la linguistique à se saisir du singulier, dont le traitement s'est souvent trouvé relégué aux marges de la discipline.

Dans les différents domaines où elle a été employée (dialectologie, sociolinguistique, linguistique énonciative, analyse de discours, sémantique des textes), en dépit de la diversité des approches dont elle était l'objet, la notion d'idiolecte a toujours manifesté une certaine stabilité définitionnelle, ce qui ne l'a d'ailleurs nullement prémunie contre le flou. Certes, comme le soulignait William Labov, son introduction par Bernard Bloch à la fin des années quarante s'est faite sur des bases qui la rendaient difficilement opératoire :

À la recherche d'un objet homogène conforme aux besoins et aux hypothèses du modèle saussurien, les linguistes n'ont cessé de rétrécir leur champ. C'est ainsi que Bloch (1948) introduisit le terme « idiolecte » pour désigner le discours d'une personne qui parle d'une seule chose à une même personne pendant un temps limité. Bien que ce terme ait été largement adopté, on peut douter que quiconque ait trouvé dans l'« idiolecte » l'homogénéité que Bloch en attendait. De toute façon, il convient de remarquer que l'existence même d'un tel concept, désignant l'objet propre de la description linguistique, représente une défaite de la notion saussurienne de *langue*, objet d'un accord social uniforme.

W. Labov, *Sociolinguistique*, Éd. de Mimit, 1976.

كلم شخصي Idiolecte [تحليل الخطاب، اللائحة الصبغة]

[الكلمة الفرنسية مأخوذة من الإنجليزية *idiolect* المشتقة انطلاقاً من اليونانية *idios* "خاص" والفرنسية *dialecte* (كلام دارج)]. يستعمل مفهوم الكلم الشخصي لتعيين استعمال لسان موسوم بنوعيات خاصة بالمتكلم. ويكثر دقته، نسبي كَمَا شخصيات المعايير والانظامات الفردية التي تكون هذه النوعيات شاهداً عليها.

يثير ضعف تطور مفهوم الكلم الشخصي في علوم اللغة ودوام الإشكالية التي يطرحها بصفة جليلة الصعوبة التي تجدها اللسانيات في ضبط المفرد الذي كثيراً ما يقع إبعاد معالجته إلى هوامس المجال العلمي.

لقد كان لمفهوم الكلم الشخصي دائماً نوعٌ من الاستقرار التعريفي في مختلف المجالات التي استعمل فيها (دراسة اللهجات، اللسانيات الاجتماعية، اللسانيات التطبيقية، تحليل الخطاب، اللائحة التصويص) وذلك رغم اختلاف المقاربات التي عالجه، وهو ما لم يحسم البتة مع ذلك من الضبابية صحيح أنه قد أدخله، في نهاية الأربعينات، برنار بلوك مثلاً أكده ويليام لايوف، على أسس تجعله صعب الاستعمال :

ما فتئ اللسانيون يحرصون مجاهم بحثاً عن موضوع متناسق يتوافق مع حاجيات السؤال السوسيري وفرضياته. هكذا أدخل بلوك (1948) مصطلح "كلم شخصي" لتعيين خطاب شخص يتحدث عن شيء واحد مع نفس الشخص أثناء مدة محدّدة. ورغم أنه وقع تبني هذا المصطلح بشكل واسع، فإنه يمكننا أن نشك أن يكون قد وجد أيّ كان في "الكلم الشخصي" التماسق الذي ينتظره بلوك من هذا المصطلح. وبهما يكن من أمر يحسن ملاحظة أن وجود هذا التصور ذاته لتعيين موضوع خاص بالوصف اللساني يشكل هزيمة لمفهوم اللسان السوسيري الذي هو موضوع اتفاق اجتماعي متماثل.

Éditions de *و لايوف، اللسانيات الاجتماعية*,
1976 - Mimit

Manuel Célio Conceição,
FCHS – Universidade do Algarve,
Faro,
Portugal,
mconcei@ualg.pt.

Lexicologie et terminologie face aux défis de la communication multilingue

Résumé

La recherche lexicologique/terminologique se fait, souvent, à partir de corpus validés et de qualité attestée. Considérer des corpus constitués par des éléments extraits de situations de communication multilingue questionne tout d'abord les critères de qualité de ces éléments et leur rapport représentatif avec le système, pose ensuite le défi de la gestion de la diversité et de la fixation/harmonisation interne à chaque langue, et crée, enfin, parallèlement, des atouts issus d'évidences de la cognition située et du parler plurilingue.

Mots-clés: communication multilingue ; diversité linguistique ; lingua franca.

La lexicologie et la terminologie bilingues ou multilingues étudient les unités lexicales (terminologiques ou non) du point de vue des rapports (cognitifs, linguistiques, communicatifs) qui s'établissent entre chaque ensemble d'unités – unités, des différentes langues, actualisées en discours et qui ont des référents communs, qui correspondent à des concepts proches ou qui dénomment les mêmes concepts.

De l'analyse de ces données ressortent, parmi d'autres résultats, des réflexions sur les possibilités de verbaliser le « découpage le réel » dans les différentes langues. À partir de l'étude des équivalents entre les langues ou des traductions sont conçus des prototypes de glossaires et des travaux lexicographiques de différentes typologies ou des études sur les conceptologies et les ontologies, par exemple. Différentes approches méthodologiques sont donc envisageables pour le travail sur des données multilingues.

L'histoire des disciplines qui étudient le lexique des différentes langues est très riche en travaux bilingues ou multilingues, néanmoins, depuis peu, de nouveaux défis leur sont posés. Les matériaux d'observation et d'analyse ont toujours été des unités (en discours ou pas) appartenant à chacune des langues qui sont analysées soit au niveau du système linguistique (« la langue », en termes de Saussure), soit au niveau de la « parole », du discours. Dans le premier cas, leur existence dictionnaire justifie les options ; dans le second cas, leur actualisation est validée par la correction des discours dans lesquels elles sont insérées, par le niveau

de connaissances reconnu aux locuteurs qui les ont verbalisées et/ou par d'autres caractéristiques éditologiques et communicatives. Peu nombreux sont les travaux lexicographiques/terminologiques multilingues qui n'ont pas comme source des textes ou des corpus validés par des locuteurs ou des spécialistes ayant un niveau de compétence linguistique qui est un des garants de la qualité du travail final.

Le grand défi actuel de la lexicologie et de la terminologie multilingues se pose au niveau de leur responsabilité en ce qui concerne l'accompagnement des langues en temps réel et donc au niveau de la constatation de l'immanence de la langue dans la communication multilingue, en général, et dans la communication professionnelle et/ou spécialisée multilingue *in vivo*, en particulier.

S'il est vrai que le multilinguisme n'est pas une situation nouvelle propre à nos temps, puisque, on pourrait dire qu'il date des temps de Babel, la communication professionnelle vit aujourd'hui de l'immédiateté et dans un désordre apparent dû à la présence de locuteurs de différentes origines linguistiques et culturelles et à un espace communicatif planétaire. En conséquence de la mobilité des locuteurs et des équipes professionnelles, de la circulation globale des connaissances, de l'utilisation des technologies de la communication qui éliminent les distances physiques, et du travail collaboratif à distance, il y a de plus en plus de situations de communication où soit une seule langue est utilisée/privilégiée (même si elle est n'est pas une des premières langues de la plupart des intervenants), soit plusieurs langues sont utilisées de manière simultanée par les intervenants. Les équipes de travail dans la plupart des secteurs productifs ne sont que rarement monolingues d'origine, même si, dans la plupart des cas, le monolinguisme leur est imposé. La mise en présence de ces locuteurs (producteurs/récepteurs discursifs) se fait soit grâce à la mobilité, caractéristique de nos temps, soit grâce aux technologies de la communication et de l'information qui nous permettent de partager nos activités en temps réel avec des partenaires où qu'ils soient dans le monde. La mondialisation ou la dimension mondiale d'une réalité, au sens de Beaud (1999), facilite la circulation des concepts, des dénominations, c'est-à-dire des façons de voir le monde et de leur verbalisation respective. En même temps, la médiatisation et l'immédiateté de la communication n'admettent de temps de recul ni par rapport aux contenus de ces communications ni par rapport à leurs pratiques. Les concepts et les dénominations voyagent sans frontières. Si ce voyage, à première vue, facilite la circulation des savoirs, il provoque le contact de langues et, partant, la glossophagie qui se traduit, par exemple, par la perte de domaines ou la perte de spécificités dénominatives, sémantiques et culturelles. Le résultat de cette glossophagie peut être le monolinguisme qui procure, sans doute, un bénéfice « immédiatiste » mais réducteur.

L'inclusion des discours multilingues (oraux et/ou écrits) dans des corpus établis à des fins lexicologiques/terminologiques oblige à une réflexion préalable sur le statut des unités qui les constituent et à redéfinir leur rapport avec les systèmes linguistiques respectifs. Les critères de correction sont, éventuellement, mis de côté pour laisser la place à ceux de la communication (en particulier celui de l'efficacité). Étant

habitué à travailler à partir de corpus bien délimités, dont les éléments constitutifs ont été validés, voire corrigés, on pourra discuter la pertinence de ce type de textes pour les études du lexique. L'argument est simple. Avec cette approche, le procédé de communication du contenu du message est pris en compte pour la construction de sa signification. Il convient aussi de préciser que dans certains domaines d'activité, dits de pointe, ce type de discours est le seul qui existe et que c'est aussi celui qui recèle les terminologies les plus actuelles. Le grand avantage du travail sur ce type de corpus est qu'il permet d'assister à la naissance de dénominations et à la déroute significative/syntaxique/pragmatique de certaines autres. Autrement dit, il permet d'assister au processus cognitif qui engendre la verbalisation et la construction discursive en situation de communication effective.

Les situations de communication professionnelle multilingue posent des défis à la lexicologie et à la terminologie que ce soit au niveau théorique ou au niveau de l'analyse linguistique. Un des plus grands défis est la révision du concept de langue vu dans le cadre de ce milieu de coexistence où plusieurs langues sont en scène dans la même situation de communication ou dans le même discours individuel. Différents scénarios de communication sont concevables et chacun aura des implications particulières dans la gestion et l'usage des langues et posera des défis spécifiques au travail sur les vocabulaires.

Parmi beaucoup d'autres, les possibilités suivantes sont envisageables.

Dans certaines situations professionnelles, plusieurs langues sont en contact, bien qu'elles y aient différents statuts, et l'une d'entre elles acquiert le statut de *lingua franca*. Cette langue peut être la langue la plus utilisée (voire la seule), bien qu'elle ne soit la première langue d'aucun des locuteurs. Un certain statut de *lingua franca* se met en place et toute verbalisation se voit, donc, filtrée par le niveau des connaissances linguistiques des intervenants dans cette langue. À titre d'exemple, et pour ne pas citer toujours le cas de l'anglais, l'espagnol assume souvent ce statut entre les étudiants Erasmus qui arrivent dans les universités portugaises en provenance de divers pays d'Europe.

Dans un groupe ou une multinationale avec différents pôles géographiquement éloignés, chaque pôle existe en une ou plusieurs langues, mais pour la communication entre eux il y a une seule langue de communication. Dans ce cas, la diversité linguistique est, sans doute, un atout pour le développement local de l'entreprise mais l'unification linguistique pour la communication interne supranationale représente un appauvrissement conceptuel puisqu'on réduit à une *lingua franca* la représentation des connaissances, des savoirs et des pratiques.

Ces situations de communication sont, en quelque sorte, des « babélisations ». Les locuteurs parlent entre eux en utilisant différentes langues, ce qui met aussi en jeu différents niveaux de compétence dans chacune de ces langues. Elles laissent entrevoir les degrés de variations en ce qui concerne les unités d'étude pour les lexicologues et les terminologues, rappelant que de nouveaux défis se posent à leur

tâche qui consiste, *grosso modo*, à identifier, analyser et gérer les différentes unités lexicales dans les différentes langues. Ces défis sont :

- Comment identifier les unités propres à une langue ?
- Comment décrire les fonctionnements linguistique, cognitif et communicatif de ces unités si elles sont attestées dans des textes qui ne respectent pas forcément les règles du système linguistique auquel elles appartiennent ?
- Comment traiter les emprunts occasionnels et les emprunts systématiques entre les langues ?
- Comment établir les équivalences dénominatives, ou significatives, entre les langues garantissant qu'il n'y a pas, par exemple, perte de domaine dans une ou plusieurs langues ?

En somme, le grand défi que pose la communication multilingue est donc l'étude détaillée des unités lexicales présentes dans ces situations d'interaction communicative, que nous décrivons ci-dessous.

Dans certains domaines d'activité internationale, tel celui de l'économie, il y a contamination des dénominations entre les langues (en particulier, évidemment, à partir de l'anglais), non seulement au niveau des emprunts lexicaux, mais aussi au niveau des systèmes morphologiques puisque certains formants passent aussi d'un système linguistique aux autres. Ayant analysé (Conceição 2009) un corpus de presse tiré de plusieurs journaux de grande circulation nationale en plusieurs langues européennes (domaine de l'économie), publiés entre le 3 et le 5 avril 2009, nous constatons que les dénominations sont soit les mêmes, soit proches dans les différentes langues. Des contextes et des concordances de termes du domaine de l'économie ont été saisis de textes des journaux suivants : *Público* (portugais), *Le Figaro* et *Le Monde* (français), *El País* et *El Mundo* (espagnol) ; *Corriere della sera* (italien), *Gandul* et *Financiarul* (roumain). Exemples de dénominations extraites (tableau 1) :

| ES | FR | IT | RO | PT |
|--|------------------------|-----------------------|---|---|
| | Actifs toxiques | <i>Titoli tossici</i> | | <i>Activos tóxicos</i> |
| <i>Autorregulación;</i> <i>Regulación</i> | Régulation des agences | | | <i>Desregulação;</i> <i>Regulação do mercado</i> |
| <i>Hedge funds</i> | <i>Hedge funds</i> | <i>Hedge funds</i> | <i>Fondurilor speculative</i> <i>(hedge funds)</i> | <i>Hedge funds</i> |
| <i>Inyecciones de dinero</i> | | | | <i>Injeção de capitais</i> |
| <i>Subprime</i> | <i>subprime</i> | <i>Subprime</i> | <i>subprime</i> | <i>subprime</i> |
| | | | <i>Lista alba</i> | <i>Lista branca</i> |
| | | | <i>Lista gris</i> | <i>Lista cinzenta</i> |

| | | | | |
|---|------------------------------------|--|--|---|
| | | | <i>Lista neagra</i> | <i>Lista negra</i> |
| | | | <i>Architecturii financiare mondiale</i> | <i>Arquitectura fianceira mundial</i> |
| <i>Paraísos fiscales</i> | Paradis fiscaux | | <i>Paradisuri fiscal</i> | <i>Paraíso fiscal</i> |
| | Droits de tirage spéciaux (DTS) | | <i>Drept special de tragere (DST)</i> | |
| <i>Supervisión macroprudencial</i> | | <i>Supervisione macroprudenziale</i> | | |
| <i>Ejercicio de risk-management</i> | | <i>Risk management</i> | | <i>Exercício de gestão do risco</i> |
| | | | <i>Politicile expansioniste</i> | <i>Políticas expansionista</i> |
| | règle du upstick | | | <i>regra upstick</i> |
| | | <i>Short selling</i> | | <i>Short selling ou venda de curtas; Vender curto</i> |

Tableau 1

Nous ne fournissons que les dénominations trouvées dans le corpus créé. Le fait de ne pas remplir tout le tableau ne veut pas dire que ces dénominations n'existent pas dans la langue considérée mais tout simplement qu'il n'y avait pas d'occurrence dans le corpus d'extraction. Ces exemples montrent que si ce type de textes est pris en compte lors de la constitution d'un corpus à des fins lexicographiques/terminologiques, il faudra bien constater que la plupart des unités ont une forme presque internationale (et un signifié linguistique aussi, peut-être) bien que le concept ne soit pas « normalisé ». Les traits conceptuels constitutifs actualisés en communication peuvent donc être conditionnés par le lecteur ou le producteur du discours. Même si ces formes semblent proches des incoterms du commerce international, leur statut, totalement linguistique, se maintient toutefois.

Étant donné la contamination qui s'observe entre les langues, du point de vue formel et du point de vue sémantique, il est vrai que les réseaux d'équivalences (pour la traduction) semblent être plus faciles à établir. Pour ce qui est de la description lexicographique/terminologique de ces unités, cette facilité n'est que trompeuse. Il faut délimiter des signifiés dans un système linguistique au sein duquel les signifiants soit n'existent pas, soit existent déjà avec un autre signifié. Il faut aussi créer des procédures d'accès au sens plein de la construction métaphorique de certaines dénominations qui s'est établie dans un autre système linguistique/culturel/référentiel. Ici, lorsque l'attestation discursive de corpus ne suffit pas apparaît le besoin du « travail de terrain » pour la quête d'informations dans le discours oral spontané ou la néologicit  est foisonnante.

Un deuxième cas de figure est la communication au sein de groupes de recherche internationaux. Par exemple celui de la communication entre les membres d'un réseau européen qui mène une recherche conjointe sur les langues et dans lequel participent près de soixante-dix universités de toute l'Europe et où la langue de travail est l'anglais. Comme nous l'avons avancé ci-dessus, le fait d'utiliser une langue étrangère ou seconde filtre la volonté de communication et le transfert de connaissances mais analysons les cas où les mêmes dénominations sont utilisées et connues par tous les locuteurs. C'est le cas de dénominations (apparemment transparentes) telles que: */employability/*, */evaluation/*, */assessment/*, */creditation/*, */skill/*, */transferable skill/*, */competence/*, */tutorial/*, */learning outcomes/* ou */contact time/*, pour lesquelles il y aura au moins autant de définitions que de locuteurs de chacune des langues en présence. On remarque que même si les dénominations sont les mêmes, vu que chacun entretient un rapport différent avec l'anglais, il y a des emprunts de signification (ou de conception) de la langue de chacun qui compliquent la communication puisque le sens des mots utilisés (en anglais) n'est pas le même pour tous.

Exemples: l'*/employability/* attribuée à un diplôme dans certains pays compte seulement les étudiants diplômés qui ont eu un emploi dans le domaine spécifique de formation et dans d'autre pays tous ceux qui ont trouvé un travail (dans n'importe quel domaine). */Evaluation/* et */assessment/* correspondent, dans certaines langues, comme le portugais, à une seule dénomination */avaliação/*. Bien que des dénominations équivalentes à */skill/* (*/capacité/*) existent, on juge souvent que la correspondance n'est pas totale et le terme anglais s'impose; néanmoins, un Portugais utilisera */transferable skill/* plutôt en tant qu'équivalent de */competência de transferência/*.

Ces termes sont très fréquents dans toute la communication au sein du projet cité, mais le signifié que chaque participant actualise en communication (en tant qu'émetteur ou en tant que récepteur) est bien différent de celui qui est attendu dans le système linguistique anglais. Les particularités de chaque acception, chaque signifié, se voient par l'analyse de discours en situation, suivant les paraphrases, les reformulations, la « mise en scène discursive ». Est-il alors utile de constituer un corpus (ou une banque de données textuelle, selon les objectifs) avec des discours issus des réunions des participants à ce projet? Si l'on entend *corpus* au sens traditionnel d'« ensemble de textes travaillés avec des objectifs définis et qui se veut de qualité et représentatif », la réponse est probablement non. Si on le considère comme représentatif de l'état concret des langues et de la communication, la réponse sera oui.

Ayant considéré ces exemples de communication multilingue, revenons aux travaux lexicologiques et terminologiques, considérant qu'ils doivent accompagner les situations réelles d'usage des langues et que les discours produits dans ces situations servent de corpus pour le dépouillement d'unités d'analyse.

Le fait d'utiliser des corpus de discours multilingues est un avantage puisque l'analyse contrastive peut mettre en évidence les processus de conceptualisation et se constitue comme possibilité de saisie de la cognition située, c'est-à-dire comme atout cognitif, communicatif et stratégique. Dans cette communication multilingue, on assiste au « parler plurilingue », tel que l'appelle Georges Lüdi (2007), et son analyse trouvera des arguments scientifiques pour montrer que la diversité linguistique est un atout pour la circulation et la mise en œuvre des connaissances dans la diversité de contextes. Le processus de conceptualisation, la cognition située, est visible par l'activation de répertoires linguistiques de différentes langues et ce caractère multilingue a un rôle de caléidoscope par rapport aux connaissances, aux savoirs et aux pratiques qu'il faut dénommer. Ici, s'annonce, donc, le grand défi, la gestion de cette nouvelle procédure de création néologique.

Il nous faut un nouveau concept de néologie puisqu'il y a de nouveaux néologismes. Au-delà du constat du caractère néologique de certaines unités, il faudra revoir les typologies de néologismes et observer leurs implications dans les systèmes ontologiques, cognitifs, dénommatifs, discursifs dans lesquels ils rentrent. La valeur sémantique d'une unité change en se mélangeant avec des sèmes qui viennent des connaissances préalables dans une autre langue. Cette nouvelle vision des unités lexicales/terminologiques fait éclater leur statut d'éléments d'un système linguistique pour les transformer en unités de communication indépendantes d'une langue ou, à tout le moins, des règles de distribution (cooccurrence et/ou restriction de sélection) dans cette même langue. Il faut, alors, une étude situationnelle du vocabulaire utilisé et une linguistique pragmatique qui montrent comment ces unités sont liées à la situation de communication, suivant, peut-être des modèles de la pragmatographie. Il s'agit donc d'un nouveau statut de l'emprunt au sein d'un concept de langue plus vaste que le concept saussurien. La langue est vue dans une conception dynamique qui prend en compte l'intercompréhension et le rapport ouvert aux autres langues. La recherche en langues et son épistémologie peuvent, sous cet angle de la communication, mettre en évidence la diversité linguistique, en tant que défi salutaire, preuve de biodiversité et enjeu de construction identitaire.

Les réponses aux défis que pose la communication multilingue à la lexicologie et à la terminologie et les méthodologies qui en découlent, dans le cadre d'analyses situationnelles, évitent la mort des langues et prouvent que le monolinguisme est loin d'être un avantage. L'étude des différentes occurrences d'unités du lexique de plusieurs langues en coexistence dans une situation de communication professionnelle et/ou spécialisée montrera aussi les procédés de circulation et de construction des connaissances et de l'information dans les discours, en contexte multilingue. Le besoin d'intercompréhension oblige, en communication multilingue, à des processus de négociation du sens dans la conversation. La saisie de ces procédés de gestion du savoir en action et des enjeux cognitifs du multilinguisme en communication seront très utiles pour la redéfinition des relations entre traits conceptuels et dénommatifs. Ces aspects intéressent la conception de banques de connaissances ou d'ontologies puisqu'ils mettent en évidence des réseaux lexicaux sémantiques et cognitifs.

Les situations de communication multilingue sont, sans aucun doute, des lieux passeurs de savoirs, où se mettent en scène des diversités que posent des défis de reconceptualisation des méthodologies et des approches en lexicologie, en terminologie et en sciences du langage et de la communication, pour ne pas réduire la communication à une *lingua franca* et la recherche en lexicologie et en terminologie au monolinguisme illusoire.

Bibliographie

- Beaud (M.), 1999 : « Présentation », dans Beaud *et al.*, coord., *Mondialisation. Les mots et les choses*, Paris, Karthala, p. 7-14.
- Conceição (M.), 2009 : « Comunicação especializada em economia internacional: multilinguismo ou língua franca », dans *Actes de la Journée Scientifique de REALITER – Terminologie et plurilinguisme dans l'économie internationale*, Università Cattolica, Milano, <http://www.realiter.net/spip.php?article1714>.
- Lüdi (G.), 2007 : « Le parler plurilingue : une catégorie *étique* ou *émique* ? », dans Gajo (L.) *et al.*, éd., *Langues en contexte et en contact. Hommage à Cecilia Serra*, Cahiers de l'ILSL 23, p. 55-63.

*Alexandra Feldekircher Müller,
Maria da Graça Krieger,
PPGLA – UNISINOS,
Canoas,
Porto Alegre,
Brésil,
alexandra.f.m@gmail.com,
mgkrieger@terra.com.br.*

*Cleci Regina Bevilacqua,
PPGLET – UFRGS,
Porto Alegre,
Brésil,
cleci.bevilacqua@terra.com.br.*

Reconnaissance de la terminologie du droit de la protection sociale dans l'Arrêt juridique sur la protection sociale

Résumé

Ce travail vise à présenter les résultats d'une recherche sur la reconnaissance de la terminologie du droit de la protection sociale en vigueur au Brésil, tout en examinant certains aspects liés à la sélection et à la reconnaissance des termes du langage de ce domaine de connaissance. Parce qu'il aborde des thématiques de la protection sociale, ce langage est véhiculé par divers supports textuels (lois, journaux, revues) et il est largement répandu dans les différentes couches de la société. Les données obtenues ont mis l'accent sur la présence de plusieurs terminologies juridiques (droit, droit de la protection sociale et autres domaines) et permis d'identifier le profil de l'utilisateur de la terminologie de la protection sociale dans les catégories récepteur et producteur, sur la base du rôle de chaque usager par rapport au texte. Il ressort de cette étude que de tels résultats sont fondamentaux pour une proposition communicative de travail dont l'objectif est de produire un glossaire sur la terminologie du droit de la protection sociale.

Mots-clés: terminologie; droit de la protection sociale; terme; usager.

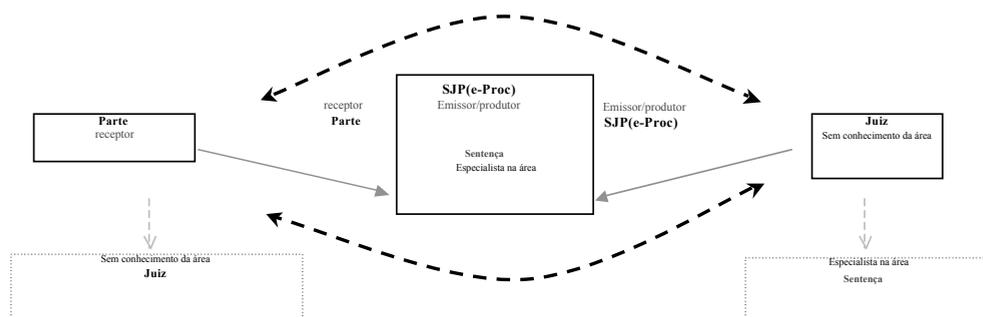
Introduction

Parce qu'il aborde des thématiques de la protection sociale, le langage du droit brésilien de la protection sociale apparaît dans différents textes (lois, journaux, revues, etc.) et est très répandu dans les différentes couches sociales de la société. Il s'agit d'un domaine de la connaissance socialement important – en termes

de bénéfices – pour une grande partie de la population brésilienne. Au vu de la normativité du droit, et plus particulièrement ici du droit de la protection sociale, les textes spécialisés sont nombreux : courriers, requêtes, actes, résolutions, arrêts, lettres rogatoires, et autres.

L'Arrêt juridique sur la protection sociale est l'un de ces documents. Sa valeur normative, légale et informative est plus dense que les autres documents dans la mesure où il vise à résoudre un problème (litige) porté à la connaissance du pouvoir judiciaire et qu'il se base sur des lois spécifiques¹. Intégrant le procès juridique, ce document est produit dans la sphère de la magistrature fédérale de la protection sociale par un spécialiste en la matière, le juge fédéral. Il s'occupe notamment de thèmes tels que : la concession ou le rétablissement des bénéfices du système de protection sociale, la révision du montant des bénéfices déjà octroyés, la reconnaissance des années de travail en milieu rural, entre autres.

Ces caractéristiques nous permettent de dire que le langage du droit de la protection sociale circule largement au sein de la population via l'arrêt juridique, remplissant un rôle à la fois informatif et social. Présentée ci-dessous (Figure 1), la situation communicative de ce texte implique deux interlocuteurs/usagers distincts : le juge fédéral – spécialiste du domaine et producteur du texte – et la partie requérante – non-spécialiste et réceptrice du texte.



Légende:

SJP (e-Proc) = *Arrêt juridique sur la protection sociale (e-Proc)*

Parte = *Partie requérante* – Juiz = *Juge*

Sem conhecimento da área = *Sans connaissances sur le domaine*

Especialista da área = *Spécialiste du domaine* – Emissor/produztor = *Émetteur/producteur*

Figure 1 : Situation communicative

Parce qu'il fait appel à un langage hautement spécialisé, qu'il comporte nombre de termes techniques et s'adresse à des interlocuteurs différents, l'arrêt juridique présente un problème communicationnel au niveau même de sa fonction : la communication d'une résolution judiciaire. Ce problème porte surtout préjudice

1 L'arrêt juridique se distingue des autres textes parce qu'il présente la décision du juge; en effet, la plupart des autres textes sont utilisés pour discuter du problème et fournir des données suffisantes pour clarifier et garantir la meilleure prise de décision.

à la partie requérante, qui comprend difficilement le texte et, conséquemment, la résolution du juge; une situation susceptible d'entraîner des difficultés légales vu la normativité du texte.

Devant cet état de fait et l'inexistence d'une mesure formelle permettant un accès plus facile au langage de l'arrêt juridique sur son espace de diffusion (le site internet de la magistrature fédérale de la protection sociale) – tel un glossaire *on-line* –, il apparaît fondamental d'identifier la terminologie spécialisée du droit de la protection sociale pour ensuite constituer un glossaire à destination des usagers connaissant peu cette terminologie.

Ainsi, l'objectif de ce travail est de présenter les résultats de la recherche² d'identification de la terminologie du droit de la protection sociale contenue dans les arrêts juridiques du système procédural électronique (*e-Proc*)³, dans le cadre de la magistrature fédérale de la protection sociale.

1 Le droit de la protection sociale et la théorie communicative de la terminologie

Le droit est un système régulateur de la société, et en tant que tel il a besoin de la langue pour subsister. Il n'existe pas de langue spécifique sur le droit et l'établissement des normes, mais une utilisation de la langue naturelle qui soutient la communication dans tous les domaines de la connaissance humaine d'une communauté linguistique donnée.

Développée au cours des dernières années, la théorie communicative de la terminologie (TCT) (Cabré 1993, 1999, 2005) questionne certains des principes de la théorie générale de la terminologie (TGT) de Wüster (1998) et se base sur le fait que la langue est unique: il n'y a pas de langue pour le droit par exemple, mais un langage du droit. Il ne s'agit ni plus ni moins que d'une utilisation de la langue naturelle avec des principes et des particularités spécifiques du domaine de la connaissance et de son type de communication spécialisé. Dans son étude sur le langage juridique, Maciel (2001a, 2001b) postule que les particularités se construisent à partir du fonctionnement de cette science et de son objectif, à savoir *dicter les lois, ordonner et faire valoir l'autorité légale*.

Ce travail considère également comme essentielles les théories du texte pour le travail terminologique envisagé dans une perspective communicative d'investigation. Partant de là, nous pouvons dire que la terminologie comporte un volet linguistique et se sert des contextes d'occurrence des termes pour analyser leur comportement et leur genèse (Krieger 2008, à paraître). Dans ce cas, les termes sont des unités

2 Cette recherche a été menée dans le cadre du master en Lettres, filière « Lexicographie et terminologie » de l'Université fédérale du Rio Grande do Sul (UFRGS), Porto Alegre, Brésil. Master achevé en 2009.

3 *e-Proc*: système procédural totalement virtuel qui permet d'accéder instantanément aux données du procès. En plus d'une réduction significative des frais, il permet une plus grande productivité et contourne la bureaucratie. L'accès se fait indépendamment de la localisation géographique.

linguistiques naturelles ayant pour caractéristique d'être des unités de la connaissance dont la valeur est spécialisée par l'utilisation (Krieger et Finatto 2004 : 78). Ils possèdent une fonction conceptuelle, dénomminative et communicative de par le contexte dans lequel ils s'insèrent (Cabré 1993 : 222).

La théorie communicative de la terminologie valorise le contexte discursif des communications spécialisées, ainsi que le rôle de la textualité et de la discursivité. Elle encourage les recherches sur les langages spécialisés et s'intéresse aux structures et aux typologies des textes produits dans les communications spécialisées. Étudier la terminologie du droit de la protection sociale revient donc à étudier la langue naturelle dans un contexte spécialisé de la connaissance ; son utilisation en fait un langage spécifique du domaine, autrement dit un langage spécialisé.

Krieger et Finatto (2004) vont dans le même sens. Dans *Introdução à Terminologia* [*Introduction à la terminologie*], elles soulignent l'importance de l'étude des termes dans les textes et l'identification des propriétés des textes dans la caractérisation des termes. Selon ces auteurs, pour la théorie communicative de la terminologie l'étude du texte est directement liée à son principe communicationnel ; cela permet d'affirmer que le texte est un *habitat* naturel des terminologies et qu'il faut le concevoir comme un objet de communication entre destinataire et destinataire. Et d'ajouter :

L'approche textuelle permet notamment d'identifier les facteurs pragmatiques de la communication spécialisée qui activent le volet terminologique assumé par différentes unités lexicales dans le contexte des communications⁴. (Krieger et Finatto 2004 :109)

Les facteurs pragmatiques peuvent être l'objet de communication, le destinataire, le destinataire, le thème, le champ du savoir... Ainsi, c'est en connaissant ces données que nous pouvons connaître ou reconnaître le lexique terminologique dans un texte spécialisé.

Pour Hoffmann (1998), le texte spécialisé est le résultat d'un acte communicatif spécialisé, composé d'une série d'items lexicaux avec des principes linguistiques d'organisation correspondant à une réalité objective donnée :

Le texte spécialisé est l'instrument ou le résultat d'une activité communicative socioproductive spécialisée. Il compose une unité structurelle et fonctionnelle (un tout) et est formé d'un ensemble ordonné et fini de propositions cohérentes pragmatiquement, syntaxiquement et sémantiquement, ou d'unités ayant valeur de proposition et qui, comme signes linguistiques complexes d'énoncés complexes de la connaissance humaine et de circonstances complexes, correspondent à la réalité objective⁵. (Hoffmann 1998 : 77)

4 Traduit par nous.

5 Traduit par nous.

D'une manière générale, les chercheurs en la matière envisagent le texte comme une représentation de la matérialisation de la communication spécialisée, dont les facteurs déterminants sont la thématique (critère thématique) et les participants de l'acte communicatif (critère pragmatique). Du fait de leur variabilité, ces deux critères (thématique et pragmatique) autorisent des typologies textuelles sur le texte spécialisé, comme le signale Hoffmann (1998).

Cette typologie peut être associée au degré de spécialisation – plus ou moins élevé – que comportent les communications spécialisées, suivant le thème et l'utilisation de la terminologie du domaine en question. Cette variation typologique est par exemple liée à la distinction entre un livre de droit (doctrine juridique) et un article de journal; chacun des deux textes emploie des critères spécifiques, car « les buts distincts de chaque type de communication et les différents destinataires prévus expliquent la variation de la densité informative, mais aussi l'utilisation plus ou moins grande de terminologies » (Krieger et Finatto 2004 : 116).

Toutefois, ces données ne sont pas valides dans la pratique textuelle de l'arrêt juridique sur la protection sociale. En effet, nous sommes en présence d'un texte hautement technique et dont le but est d'annoncer un arrêté judiciaire à un destinataire profane. L'utilisation de l'appel à la « variation de la densité informative » est inexistante, de même que l'utilisation de l'appel « plus ou moins grand à l'utilisation de terminologies », qui permettrait de rendre le texte compréhensible pour son destinataire. On le voit, l'auteur du texte *ne se soucie pas* des aspects pragmatiques de la situation communicative qui se crée autour de ce texte spécialisé⁶.

Ciapuscio (1998) parle également de la variation conceptuelle qui s'établit selon la typologie textuelle de la communication spécialisée. Cette variation conceptuelle est liée aux degrés de spécialisation de certains textes au détriment d'autres, à l'image du livre et de l'article cité précédemment. Dès lors,

Il y a beaucoup de variations dénomminatives, suivant qu'il s'agit d'un texte hautement spécialisé ou destiné à une diffusion générale. De la même manière, le degré de la densité informative qui y est rattachée varie selon le type de destinataires, spécialistes ou public profane. Dans ce cas de variation, les formulations linguistiques qui, tels des connecteurs comme *autrement dit* ou *c'est-à-dire* visent à expliquer des aspects conceptuels plus complexes, tendent à être fréquentes⁷ (Krieger et Finatto 2004 : 117).

Ces observations nous permettent de mieux comprendre l'importance du texte pour les études terminologiques et la nécessité d'études linguistiques et structurelles des textes. Le texte est chaque fois plus perçu comme l'*habitat* naturel des terminologies et le *vecteur de l'articulation et du comportement* des objets de la terminologie;

6 Il est possible d'affirmer cela au vu de la proximité du langage du texte avec les autres textes juridiques (tels ceux présents dans un ouvrage de droit), qui s'adressent à des usagers spécialistes.

7 Traduit par nous.

d'autre part, une communication claire, basée sur les *partenaires communicatifs*⁸ *spécialiste/profane* est possible si l'on observe la typologie textuelle et le *modus dicendi* ou le type de traitement donné aux concepts les plus complexes.

Adoptée comme référentiel théorique parce qu'elle tient compte de l'univers discursif dans lequel sont insérés le langage spécialisé et sa terminologie, la théorie communicative de la terminologie nous permet d'identifier les termes du droit de la protection sociale dans leur contexte d'utilisation : les arrêts juridiques. De plus, les points de vue sur le texte spécialisé envisagés comme apport théorico-méthodologique se révèlent être une démarche indispensable pour l'objectif de la présente recherche.

En complément du référentiel théorico-méthodologique, nous utilisons également les apports de la linguistique de corpus pour travailler sur un ensemble de textes spécialisés. Plus spécifiquement, sur l'analyse de Pearson (1999 : 16) concernant l'utilisation des corpus textuels comme un recueil de textes produits à des fins et selon des méthodes spécifiques.

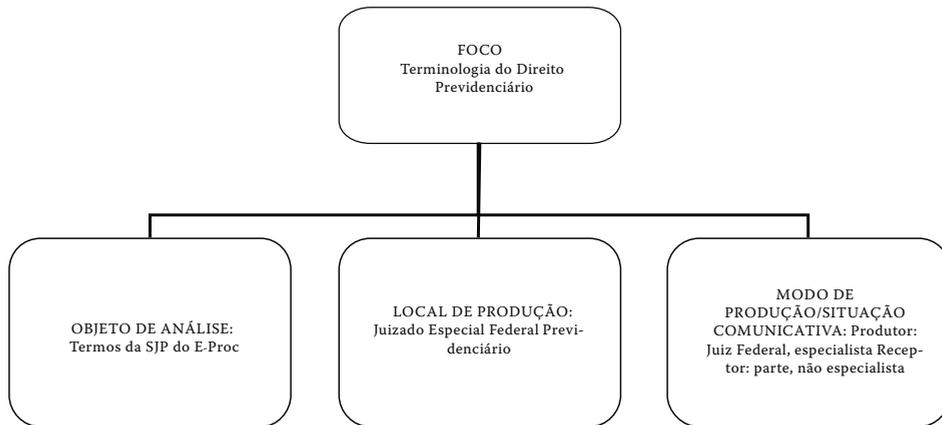
Nous pensons que l'utilisation d'un corpus informatisé susceptible d'être analysé linguistiquement est indispensable pour extraire les unités terminologiques du droit de la protection sociale ; un corpus basé sur les principes de la linguistique de corpus⁹ (Sardinha : 2004). Composé de 120 arrêts juridiques des *e-proc* recueillis en 2007 dans trois circonscriptions judiciaires de la magistrature fédérale du droit de la protection sociale (États du Rio Grande do Sul, de Santa Catarina et du Paraná, intégrant de la quatrième Région de la Justice fédérale), ce corpus traité avec les outils du logiciel *Corpógrafo* a permis de mettre en place un travail de reconnaissance de la terminologie du droit de la protection sociale.

2 Identification de la terminologie du droit de la protection sociale

Identifier et comprendre le langage du droit de la protection sociale ainsi que la terminologie contenue dans l'arrêt juridique implique de connaître leur utilisation et leur contexte de création, comme le montre le schéma ci-dessous (figure 2) :

8 Expression utilisée par Pearson (1999) pour désigner les interlocuteurs – producteur et récepteur – du texte spécialisé.

9 Certains de ces principes sont : l'origine, le but, la composition, la mise en page, la représentativité, l'étendue des textes.



Légende:

Foco: = *Objectif:*

Terminologia do direito previdenciário = *Terminologie du droit de la protection sociale*

Objeto de análise: = *Objet d'analyse:*

Termos da SJP do processo eletrônico (e-Proc) = *Termes de l'Arrêt juridique sur la protection sociale sur e-Proc*

Local de produção: = *Lieu de production:*

Juizado especial federal previdenciário JEFP = *Magistrature fédérale de la protection sociale*

Modo de produção/situação comunicativa = *Mode de production/situation communicative*

Produtor: Juiz Federal, especialista = *Producteur: juge fédéral, spécialiste*

Receptor: parte autora, não especialista = *Récepteur: partie requérante, non-spécialiste*

Figure 2 : Aspects de la situation communicative dans le cadre de l'arrêt juridique sur la protection sociale

Sur la base des informations présentées, l'étape méthodologique d'identification des termes du droit de la protection sociale a consisté à analyser le texte de l'arrêt juridique sur un plan macrostructural et microstructural. Cette analyse nous a permis de reconnaître une macrostructure récurrente dans le texte, devenue prototypique du texte étudié et du domaine discursif du droit: 3 segments distincts, à savoir introduction, développement et conclusion¹⁰.

Quant à la microstructure, elle comporte des termes dont la valeur et la structure ne sont identifiées que dans le genre arrêt du contexte juridique. C'est par exemple le cas de l'unité *Vu* placée au début du texte en forme de paragraphe et qui reprend, par inférence du lecteur, tout le contenu traité dans les étapes procédurales précédentes¹¹, se caractérisant ainsi comme *modus dicendi* du domaine.

10 Cependant, du point de vue juridique (Loi n° 9 099/95, art. 38) les arrêts juridiques des magistratures fédérales spécialisées sont formellement composés de fondements et d'un dispositif, ce qui équivaut au développement et à la conclusion; le segment introduction, identifié dans notre analyse, n'apparaît pas.

11 Étant donné que le procès juridique se compose de plusieurs étapes procédurales, donc de différents documents (plusieurs genres textuels) dans lesquels sont décidés certains points importants pour le développement du procès, l'expression *Vu*, etc. reprend ces données.

La figure suivante illustre la structure identifiée dans le texte, avec les segments macrostructural et microstructural¹²:



Légende:

Texto especializado = *Texte Spécialisé*

Macroestrutura = *Macrostructure*

Introdução = *Introduction*

Desenvolvimento = *Développement*

Conclusão = *Conclusion*

Vistos, etc. = *Vu, etc.*

Dispensado o relatório = *Dispensé de rapport*

Decido = *Je décide*

Interesse processual = *Intérêt procédural*

Ante o exposto = *Au vu de ce qui est exposé*

Julgo improcedente o pedido = *Je juge la demande non justifiée*

Microestrutura = *Microstructure*

Figure 3 : Structure du texte

Les éléments microstructuraux identifiés façonnent la macrostructure globale du texte. Pour déterminer la terminologie du droit de la protection sociale, il convient d'analyser le texte du point de vue de sa microstructure, c'est-à-dire des unités terminologiques en soi. Néanmoins, l'analyse conjointe de la macrostructure permet de saisir le fonctionnement et le rôle des unités terminologiques dans la structuration du langage au sein du texte et dans la diffusion de l'information.

¹² Van Dijk (1975) fut l'un des premiers à proposer les concepts de macrostructure et microstructure du texte. Pour l'auteur, la macrostructure est une structure de signification globale du texte, dérivée de la microstructure ou base du texte. Pour Vilela (1999), la microstructure et la macrostructure textuelles sont des unités qui se situent dans le domaine du cognitif, du sémantique: il s'agit d'une configuration de la connexion globale du texte. En somme, nous pouvons en déduire que les macrostructures textuelles permettent la formulation d'une idée globale du texte, tandis que les microstructures constituent des unités en soi et s'articulent formellement pour rendre la macrostructure cohérente.

Dans ce sens, le texte spécialisé est source d'entendement de l'utilisation et de la fonction de chaque unité terminologique. Nous présentons à titre d'exemple le texte spécialisé Arrêt juridique sur la protection sociale avec une délimitation de ses microstructures et macrostructures :



Figure 4 : Macrostructure et microstructure de l'arrêt juridique sur la protection sociale.

Chacun des trois segments est délimité par des pointillés, conformément à la macrostructure du texte, et chaque élément de la macrostructure qui apparaît en rouge souligne un aspect de la microstructure. Le texte nous permet de savoir que *Dispensado o relatório* [Dispensé de rapport] se réfère au topique du texte qui contiendrait le rapport de la demande de la partie requérante, étayée par les lois référées¹³ et se caractérisant comme une introduction dans le texte. *Vistos, etc.* [Vu, etc.] et *Dispensado o relatório* [Dispensé de rapport] – toujours dans cette disposition textuelle – font aussi partie de l'élément initial.

L'unité *Decido* [Je décide] annonce l'exposé du problème, c'est-à-dire le développement de l'action, tout en introduisant le sujet qui sera traité dans le texte avec l'aide du paragraphe suivant (*Trata-se de ação revisional de benefício previdenciário na qual o (a) s autor (a) (es) pretende (m) ver revisado o cálculo de concessão de seu benefício, [...]*¹⁴). Même s'il varie un peu du fait de l'intérêt procédural de chaque partie, ce paragraphe conserve la même construction phrasale et la même disposition dans le texte.

Les unités *Ante o exposto* [Au vu de ce qui est exposé] et *julgo parcialmente procedente* [je juge partiellement justifié] annoncent pour leur part un segment représentatif de la structuration des idées conclusives du texte et impliquent la finalisation

13 Cf. note 9.

14 Il s'agit de l'action révisoire de bénéfice de protection sociale à travers laquelle l'(les) auteur(s) demande(nt) la révision du calcul de concession de son (leur) bénéfice [...].

et la compréhension de la signification totale du texte en fonction de sa charge sémantique.

Au-delà des données sur les éléments qui caractérisent chaque partie de la macrostructure, sont présentes dans le texte plus d'une terminologie spécialisée, autrement dit un entrelacement des langages spécialisés. Dans la mesure où le droit de la protection sociale est un champ ou une branche¹⁵ du droit, les textes produits comportent, en plus des termes du droit de la protection sociale (*bénéfice, impôt prélevé sur salaire, assuré, revenu mensuel initial*, etc.), des termes de la terminologie générale du droit (*procès-verbaux, magistrature spécialisée, juridictionnelle*, etc.) ainsi que d'autres branches, telles que le droit fiscal¹⁶ et administratif. En plus des données des domaines du droit, sont incluses des terminologies d'autres domaines de la connaissance, à l'exemple de l'expression *formulaire DSS 8 030* qui provient de l'ingénierie de sécurité du travail et d'*ORTN/OTN*¹⁷, issu du secteur économique.

Ces éléments montrent le rôle que chaque unité peut exercer au sein du texte, à l'image de l'expression *Vu, etc.* déjà mentionnée précédemment. Dans un même temps, elles caractérisent la manière dont le domaine juridique et celui du droit de la protection sociale en particulier représentent les informations relatives au domaine dans la mise en forme d'un texte, se prévalant de plus d'une terminologie spécialisée¹⁸ pour communiquer.

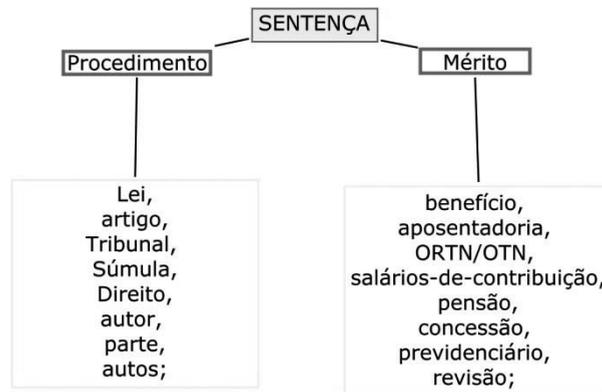
Il est possible d'aller de l'avant et de considérer les termes rencontrés comme des unités représentatives des procédures juridiques ou des contenus/actions des arrêts. Cette classification se baserait sur le paradigme procédure/mérite, le premier étant la partie fixe du texte, intégrant la structure et l'organisation du pouvoir judiciaire, et le second la partie variable, mettant en évidence les actes institués par l'arrêt de la magistrature fédérale du droit de la protection sociale. Cette possibilité de variation classificatoire est liée à la diversité même des thèmes abordés par le droit de la protection sociale. S'ensuit une visualisation de ce paradigme avec les exemples de termes :

15 D'après les spécialistes du domaine juridique, les termes *branche* et *domaine* sont souvent utilisés comme synonymes.

16 À ce sujet, Martins (2002 : 53-61) affirme : « La relation entre le droit de la sécurité sociale, plus particulièrement de la protection sociale, et le droit fiscal est très grande », à commencer par le maintien financier du système de sécurité sociale. Les termes adoptés sont : fait générateur, sujet actif, sujet passif, incidence, base du calcul, contribuables. « Le droit fiscal est une source subsidiaire du droit de la sécurité sociale, avec lequel il entretient des relations étroites et a des influences décisives en matière de maintien financier. Si les particularités des situations de la sécurité sociale sont respectées, les normes fiscales pourront être invoquées pour aider à l'application et à l'interprétation de la législation de celle-ci ».

17 ORTN : obligation réajustable du Trésor public/OTN : obligation du Trésor public.

18 Nous entendons par terminologie spécialisée l'ensemble des termes d'un domaine de connaissance donné.



Légende :

Sentença = *Arrêt* – Procedimento = *Procédure* – Merito = *Mérite*

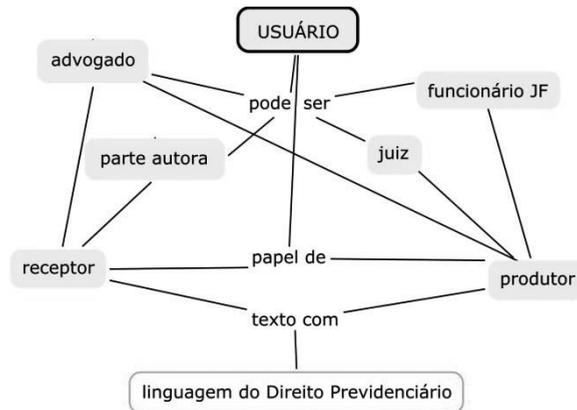
Lei, artigo, tribunal, sumula, direito, autor, parte, autos; = *Loi, article, tribunal, résumé de la décision, droit, auteur, partie, procès-verbal;*

Benefício, aposentadoria, ORTN/OTN, salários-de-contribuição, pensão, concessão, previdenciário, revisão = *Bénéfice, retraite, ORTN/OTN, impôt prélevé sur salaire, pension, concession, protection sociale, révision*

Figure 5 : Paradigme procédure/mérite

Enfin, pour bien comprendre le texte et sa terminologie, il faut connaître les caractéristiques de son usager. Nous avons procédé à son identification sur la base des rôles de récepteur et d'émetteur conjointement à l'acte communicatif instauré.

Les récepteurs du texte spécialisé et de la terminologie du droit de la protection sociale sont la partie requérante et son avocat; quant aux émetteurs, il s'agit des juges et des fonctionnaires de la magistrature fédérale du droit de la protection sociale, comme le montre le schéma ci-dessous :



Légende:

Usuario = *Usager* – Advogado = *Avocat*

Funcionario JF = *Fonctionnaire magistrature*

Pode ser = *peut être* – Receptor = *récepteur*

Papel de = *rôle de* – Juiz = *judge* – Parte autora = *partie requérante*

Produtor = *producteur* – Texto com = *texte avec*

Linguagem do direito previdenciario = *langage du droit de la protection sociale*

Figure 6 : Identification de l'usager

On a pu constater que les usagers de la terminologie du droit de la protection sociale présente dans l'arrêt juridique de l'*e-Proc* sont au nombre de quatre : la partie requérante et les avocats (récepteurs), les juges et les fonctionnaires de la magistrature fédérale (émetteurs) ; deux catégories différentes d'usagers de l'arrêt juridique.

Vu qu'ils sont les plus concernés par l'arrêt juridique, la partie requérante et le juge constituent les principaux agents de ce processus communicatif : la première parce qu'elle désire connaître l'avis de l'arrêt et le deuxième parce qu'il formule une résolution. L'avocat et les fonctionnaires représentent des agents intermédiaires de ce processus communicationnel, ils sont responsables de la mise en œuvre des étapes procédurales. En aidant la partie requérante au cours du processus et dans le cas de l'arrêt juridique, l'avocat est un récepteur du langage du droit de la protection sociale. Mais il peut aussi être à certains moments du procès un producteur de ce langage, notamment lors de l'élaboration de la requête initiale¹⁹. Quant aux fonctionnaires, ils sont chargés d'accompagner et de travailler à chaque étape procédurale, jusqu'à la prononciation de l'arrêt.

¹⁹ Ainsi, le même avocat peut remplir la double fonction de récepteur et producteur du langage. Pendant l'étape procédurale nommée arrêt, il est récepteur du langage. Il convient d'ajouter ici que le procès de la magistrature fédérale du droit de la protection sociale a pour caractéristique de ne pas nécessiter d'avocat pour traiter d'actions dont le montant est inférieur à 60 salaires minimums brésiliens.

Le schéma qui suit présente le profil des usagers dans les catégories de récepteur et de producteur de la terminologie des arrêts juridiques sur la protection sociale, ainsi que leurs principales caractéristiques :



Légende:

Usuario da terminologia do DP = *Usager de la terminologie du droit de la protection sociale*

Receptor = *récepteur* – Produtor = *producteur* – Leigo = *Profane*

Pessoa que possui baixo ou nenhum grau de escolaridade, podendo chegar ao ensino fundamental
= *personne de niveau scolaire faible ou inexistant, en deçà du niveau secondaire*

Semileigo = *Semi-profane*

Pessoa que possui um grau de instrução maior como ensino fundamental ou ensino superior, mas que não possui nenhum vínculo ou conhecimento da área especializada = *personne de niveau scolaire secondaire ou universitaire, mais connaissance ou lien avec le domaine de spécialité*

Especialista = *spécialiste*

Pessoa com alto grau de escolaridade e formação superior na área de direito; experiência e conhecimento da área = *personne de formation supérieure en droit; expérience et connaissance du domaine*

Semi-especialista = *semi-spécialiste*

Pessoa com boa escolaridade como ensino médio e/ou ensino superior incompleto em direito, e conhecimento e/ou atuação na área = *personne de niveau scolaire secondaire et/ou universitaire inachevé en droit, et connaissance et/ou travail dans le domaine de spécialité.*

Figure 7 : Profil de l'usager

De tels résultats permettent de faire avancer considérablement le travail terminologique à visée linguistique, communicative et cognitive, dont l'objectif final est la production d'un glossaire spécialisé. En effet, ils délimitent le profil de l'usager et identifient le type d'organisation et de comportement du texte de l'arrêt juridique sur la protection sociale, de même que sa terminologie.

Considérations finales

La description de la macrostructure et de la microstructure des arrêts juridiques, ainsi que l'identification de ses usagers, montrent que ce type de texte est un document qui fait partie intégrante de la procédure juridique, présenté *on-line* et élaboré par un juge fédéral. Il vise à octroyer un avis final ou semi-final sur le thème avancé par la partie requérante, favorable ou non à sa demande. Il est donc un texte

spécialisé normatif et informatif, produit par un spécialiste dans un contexte de production formelle et ordonné sémantiquement et pragmatiquement suivant le champ dans lequel il s'insère.

L'arrêt juridique a pour caractéristique une macrostructure récurrente et une microstructure susceptible d'illustrer la terminologie du domaine et son rôle vis-à-vis du texte. D'autre part, il se compose d'un entrelacement de différents langages spécialisés et permet de classer les termes à l'intérieur du paradigme procédure *versus* mérite.

Finalement, l'identification d'un usager récepteur profane, semi-profane ou spécialiste et d'un producteur spécialiste ou semi-spécialiste permet de mettre en évidence les partenaires communicatifs impliqués dans l'acte de diffusion de la connaissance spécialisée, via le texte de l'arrêt juridique sur la protection sociale.

Les données obtenues attestent notamment l'importance, pour la réalisation d'un travail terminologique, de l'analyse et de la description du contexte de production, des caractéristiques du texte et de ses usagers.

Sur la base des réflexions théoriques et des résultats présentés, la prochaine étape de notre travail consistera en l'élaboration d'un glossaire sur le droit de la protection sociale, destiné à être publié sur des sites internet de consultation procédurale dans le cadre du droit de la protection sociale. Une telle analyse justifie la nécessité d'organiser et de diffuser, à travers un glossaire structuré pour un usager non spécialiste, la terminologie du droit de la protection sociale. En réalité, il s'agit d'un instrument fondamental pour tout travailleur brésilien.

Bibliographie

- Cabré (M.T.), 2005 : *La terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*, Girona, Documenta Universitaria, 362 p.
- Cabré (M.T.), 1999 : *La terminología: representación y comunicación: elementos para una teoría de base comunicativa e otros artículos*. Barcelona, IULA, 369 p.
- Cabré (M.T.), 2005 : *La terminología*, Barcelona, Antardida/Empuries, 1993, 529 p.
- Castro (C.A.P.) et Lazzari (J.B.), 2005 : *Manual do Direito Previdenciário*. 6^e éd., São Paulo, LTR Editora.
- Ciapuscio (G.E.), 1998 : « La terminología desde el punto de vista textual: selección, tratamiento y variación », dans *Organon*, vol. 12, n^o 26, p. 43-65.
- Dijk (T. A.) Van, 1978 : *The Puerto Rico Lectures on the Structures and Function of Discourse*. Amsterdam.
- Hoffmann (L.), 1998 : *Llenguatges d'especialitat*. Barcelona, Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada.
- Krieger (M.G.), 2008 : « Por que Lexicografia e Terminologia: relações textuais? » *Anais do VIII CELSUL*. Porto Alegre, UFRGS.
- Krieger (M.G.), Finatto (M.J.), 2004 : *Introdução à Terminologia: Teoria & Prática*. São Paulo, Contexto.

- Maciel (A.M.B.), 2001a: « Terminologia, linguagem de especialidade e dicionários », dans Krieger, (M.G.), Maciel (A.M.B.), org., *Temas de Terminologia*. Porto Alegre/São Paulo, Ed. Universidade/UFRGS/Humanitas/USP.
- Maciel (A.M.B.), 2001b: *Para o reconhecimento da especificidade do termo jurídico*. Porto Alegre: UFRGS, Thèse de doctorat.
- Martins (S.P.), 2002 : *Fundamentos de Direito da Seguridade Social*. São Paulo, Atlas.
- Pearson (J.), 1998 : *Terms in context*. Amsterdam & Philadelphia, Benjamins.
- Sardinha (T.B.), 2004 : *Linguística de Corpus*. São Paulo, Manole, 410 p.
- Vilela (M.), Koch (I.V.), 1999 : *Gramática da Língua Portuguesa*. 2^o éd. Coimbra, Almedina.
- Wüster (E.), 1998 : *Introducción a la teoría general de la terminología y a la lexicografía terminológica*, Barcelona, IULA, Trad. Anne-Cécile Nokerman.

Françoise Bacquellaine,
FLUP/CLUP – Unidade n°22 da FCT,
Universidade do Porto,
Porto,
Portugal,
franba@letras.up.pt – shirleybac@gmail.com.

Technologies de pointe, terminologies et diversité linguistique : la terminologie *Bluetooth* en français et en portugais

Résumé

Le défi de la diversité se pose notamment au niveau du discours scientifique et technique, où la terminologie joue un rôle central. La terminologie de la jeune technologie *Bluetooth* constitue un objet d'étude néonymique privilégié. Certes, la comparaison des procédés de formation secondaire et de l'usage en français et en portugais confirme l'influence de l'anglais sur ces deux langues, mais elle révèle aussi la nécessité de reformuler les termes internationaux dans la langue du discours. La terminologie *Bluetooth* existe donc aussi bien en français qu'en portugais. Le seul obstacle à sa diffusion est la mobilisation des ressources nécessaires.

Mots-clés : néonymie ; technologie ; diversité linguistique ; français ; portugais.

Introduction

Le 15 octobre 2004, un groupe d'éminents spécialistes se réunissaient en colloque à Paris pour se pencher sur les néologies contemporaines. Lors de la discussion suivant l'intervention de John Humbley intitulée « Dictionnaire de néologismes : porte d'entrée des anglicismes ? Perspective européenne des dictionnaires de néologismes et des services linguistiques en ligne » (Depecker et Dubois 2005 : 27-36), Louis-Jean Rousseau (*idem* : 36-37) constate que si l'emprunt « n'est pas gênant en soi », « l'emprunt massif » peut entraîner la perte de domaines « non seulement pour la terminologie mais aussi pour la langue d'usage », comme c'est le cas en suédois, notamment dans le domaine « des technologies de l'information et de la communication » ; John Humbley (*idem* : 37) confirme que « la perte de domaines en danois » devient une préoccupation pour certains Danois ; enfin, André Le Meur (*ibidem*) suggère de vérifier « dans quelle mesure l'informatique est un domaine perdu pour l'allemand ». L'enjeu est donc la préservation de la diversité linguistique de l'élément le plus dynamique de toute langue vivante : les signes, mots de la langue courante ou termes des langues de spécialité, inventés au fur et à mesure des besoins pour désigner les nouvelles entités, les nouveaux concepts que l'homme invente jour après jour, à un rythme de plus en plus rapide.

D'autre part, dans l'introduction aux actes de ce colloque, Loïc Depecker (idem : 9-10) recommande l'« exercice passionnant » qui consiste à comparer les néologies, étant donné que « la néologie d'une langue est unique et semblable à elle-même » et que l'on peut donc considérer qu'il existe « plusieurs » néologies, « chaque langue ayant la sienne et ses propres moyens d'invention ».

Alors, qu'en est-il du français et du portugais ? Les terminologies de technologies de pointe sont-elles aussi perméables à l'anglais dans ces deux langues romanes que dans les langues germaniques ? Les procédés de francisation et de portuguisation sont-ils identiques ? Les locuteurs francophones et lusophones ont-ils les mêmes habitudes langagières ? Les résultats d'une étude récente de « néonymie comparée » (néologie en terminologie) dans le domaine de la technologie *Bluetooth* apportent quelques éléments à l'enquête.

Nous verrons d'abord comment se présente le panorama linguistique général du village global d'après le modèle gravitationnel présenté par Louis-Jean Calvet (1999, *apud* Calvet 2002 : 26-27) pour nous pencher ensuite sur les résultats de l'étude menée en 2006-2008. Par définition, la terminologie est interdisciplinaire, c'est une discipline forcément transversale. Dans ce cas, l'étude se situe au confluent de la linguistique et du génie des télécommunications, mais elle est aussi multilingue, c'est pourquoi cette étude de néonymie comparée a été dirigée par une linguiste anglophone, Belinda Maia (*Faculdade de Letras da Universidade do Porto*) et par un ingénieur lusophone, Nuno Almeida (*Faculdade de Engenharia da Universidade do Porto*), afin de mesurer l'ampleur de la menace qui pèse sur la diversité linguistique dans les terminologies de pointe en français et en portugais.

1 Les langues dans le village global

L'intensification des échanges dans le village global exerce une influence plus marquée qu'autrefois sur toutes les langues et l'on peut se demander s'il faut envisager l'avenir du patrimoine linguistique de l'humanité dans une perspective pessimiste ou plutôt optimiste, comme le font Henriette Walter (1997 et 2001) ou Édouard Glissant¹ (1997). Certes, l'anglais est la seule langue hypercentrale du modèle proposé par Calvet (2002 : 26-27). Mais dans un article intitulé « Vie et mort des langues ; les locuteurs décident », Calvet (2000) entrevoit la possibilité que l'anglais, à l'instar du latin, se diversifie au point de disparaître au profit de nouvelles variantes, ce qui correspondrait finalement à un enrichissement de la diversité.

Parmi les langues supercentrales, centrales et périphériques, deux tendances se dessinent. Certaines ont été rétablies dans la plupart de leurs fonctions, non seulement grâce à des efforts institutionnels, mais surtout grâce à la volonté des locuteurs (Calvet 2002 ; Klinkenberg 2001). D'autres meurent, faute de locuteurs, ou

1 La notion de « crise » des langues évoquée dans la communication orale a été supprimée et la référence à la notion de « Tout-Monde » d'Édouard Glissant a été ajoutée sur les conseils de Michèle Gendreau-Massaloux lors de la discussion qui a suivi la communication.

commencent à perdre des domaines du savoir, car les terminologies de technologies de pointe proviennent presque exclusivement de l'anglais, à un rythme de plus en plus rapide et les autres langues ont du mal à assimiler cet afflux de nouveaux termes ou néonymes qui « envahissent » ces domaines d'autant plus facilement que la communication scientifique internationale tend à devenir unilingue : la plupart des scientifiques sont convaincus que pour être lu, il faut écrire en anglais (voir, p. ex., Chansou 2003 : 110).

Certains s'acharnent à « défendre » les langues « minoritaires » contre « l'invasion » anglo-saxonne en adoptant une politique linguistique prescriptive ou incitative ou en « s'alliant » contre « l'ennemi » commun. D'autres voient dans cet acharnement l'expression d'un sentiment xénophobe et s'opposent à toute tentative de contrarier le cours naturel des langues vivantes (voir, p. ex., Faraco 2002). Les plus modérés admettent la nécessité d'aménager la terminologie pour permettre l'accès au savoir dans d'autres langues que l'anglais.

2 La terminologie *Bluetooth* en français et en portugais

La technologie *Bluetooth* est née au tournant du siècle. La terminologie qui permet d'en parler est un ensemble dynamique qui évolue au rythme des progrès de cette jeune technologie. Elle se forme en anglais et comporte un volume important de néonymes, ce qui favorise les emprunts massifs. Elle constitue donc un objet d'étude présentant toutes les caractéristiques de l'usage actuel en matière de terminologie d'un domaine à la pointe du progrès.

L'hypothèse selon laquelle les terminologies de pointe, telles que la terminologie *Bluetooth*, sont actuellement des domaines en danger dans les langues supercentrales, telles que le français ou le portugais, a été vérifiée selon une perspective qui se veut à la fois extralinguistique (objet ou entité), supralinguistique (concept), intralinguistique (français, portugais) et interlinguistique (français versus portugais), pour relever le défi lancé par Depecker en 2005 (2005 : 11).

2.1 Méthodologie

L'étude de l'objet observable que sont les réseaux *ad hoc* établis au moyen de la technologie *Bluetooth* a été menée avec l'aide d'un spécialiste du domaine qui m'a permis d'acquérir les connaissances suffisantes pour concevoir une partie de la structure conceptuelle de cette technologie de pointe.

Des corpus comparables ont été constitués à partir d'un document audiovisuel en portugais² et de documents écrits techniques, scientifiques et didactiques en anglais international, en portugais ibérique et en français d'Europe et du Canada.

² Le 29 mai 2007, le professeur Nuno Almeida de la *Faculdade de Engenharia da Universidade do Porto* nous a permis de filmer sa leçon d'une heure sur la technologie *Bluetooth*. Cette leçon s'adressait à des étudiants de 4^e et 5^e années en génie informatique dans le cadre de la discipline « *Comunicações móveis* ».

Les corpus écrits ont été exploités au moyen du *Corpógrafo*, un logiciel conçu par l'équipe du pôle de Porto de la *Linguateca*, sous la direction du professeur Belinda Maia. Ce logiciel a également permis d'élaborer une base de données terminologiques couvrant 69 concepts sélectionnés sur les conseils de l'ingénieur. Les données relevées dans le document audiovisuel (termes, définitions, contextes) ont été insérées manuellement.

Cet échantillon terminologique trilingue – anglais, français, portugais – comporte 486 entrées : 122 en anglais, 205 en français et 146 en portugais européen, auxquelles 13 entrées en portugais du Brésil ont été ajoutées en raison de leur intérêt néonymique. Plusieurs entrées correspondent donc à chaque concept, et ce, quelle que soit la langue. En anglais, il s'agit souvent de sigles et de leur transcription. En français et en portugais, on retrouve ces termes internationaux auxquels s'ajoutent des « reformulations » au sens large, telles que les conçoit Célio Conceição (2005).

2.2 Modèle d'analyse

Au niveau intralinguistique, l'analyse a porté sur les procédés de formation primaire des néonymes en anglais et les procédés de formation secondaire en français et en portugais. Seuls ces derniers nous intéressent aujourd'hui. Le modèle d'analyse s'inspire notamment de celui de Depecker dans *L'invention de la langue. Le choix des mots nouveaux* (2001 : 402-415). Son modèle de « procédés de traitement » des nouveaux concepts va de l'« emprunt direct » à l'« emprunt par traduction ». L'ordre de présentation des procédés va du plus anglicisé au plus francisé ou portuguisé.

L'emprunt direct peut prendre la forme d'un « emprunt intégral (importation de la forme et du sens sans modification significative) » ou d'un « emprunt aménagé (aménagement de la forme, et éventuellement du sens) ». Contrairement au calque, l'emprunt aménagé ne suppose pas de traduction. L'aménagement de l'emprunt se réalise par la francisation – ou la « portuguisation » – du suffixe ([1] *active* → *actif/activo*; [2] *member* → *membre/membro*), par une régulation morphologique ou graphique du radical ([3] *address* → *adresse/endereço*) voire par une adaptation syntactique de la forme anglaise en inversant les éléments des syntagmes et en ajoutant un ou plusieurs joncteurs si nécessaire ([4] *active member address* → *adresse de membre actif/endereço de membro activo*). La catégorie « emprunt bilingue » a été ajoutée au modèle pour rendre compte des nombreux termes composés d'un noyau aménagé en français ou en portugais et d'un « internationalisme », c'est-à-dire un sigle anglais ([5] *L2CAP* (*Logical Link Control and Adaptation Protocol*)³ → *protocole L2CAP/protocolo L2CAP*) ou un emprunt intégral à l'anglais ([6] *sniff mode* → *mode sniff*; [7] *standby state* → *estado (de) standby*).

Pour Depecker, l'emprunt par traduction consiste en un « calque (traduction littérale du terme étranger) », un « calque transpositionnel (traduction non littérale) » ou une « synthèse néologique (reconceptualisation) ». Dans le cas qui nous intéresse,

3 L'emploi des majuscules correspond à l'usage en anglais dans ce type de discours.

quatre catégories principales se dégagent: le calque bilingue ([8]), le calque unilingue ([9]), le calque transpositionnel bilingue, où la transposition se combine à un internationalisme et à un aménagement ([10]) ou à une traduction littérale ([11]) et enfin le calque transpositionnel, dont la sous-catégorie la plus proche de la « synthèse néologique » de Depecker est le transfert d'image à partir du concept ([12] et [13])⁴. Toutes ces catégories sont attestées dans le corpus en français, les calques transpositionnels en portugais des exemples [12] et [13] proviennent du corpus en portugais du Brésil:

[8] *ACL (Asynchronous Connectionless) link* → *liaison ACL/ligação ACL*

[9] *frequency hopping spread spectrum* → *étalement de spectre par sauts de fréquence/espalhamento espectral por saltos em frequência*

[10] *Bluetooth core system* → *système principal Bluetooth*

[11] *Bluetooth protocol stack* → *pile protocolaire Bluetooth/pilha protocolar Bluetooth*

[12] *sniff* → *repos actif/estado interessado* (PB)

[13] *standby* → *estado alienado* (PB)

2.3. Analyse des résultats

Pour déterminer l'importance de l'influence de l'anglais sur la terminologie *Bluetooth* dans la communication scientifique, technique et didactique en français et en portugais européen, deux comparaisons ont été établies. La première consiste en une analyse statistique de l'éventail des procédés de formation secondaire mis en œuvre dans chacune des deux langues romanes, sans tenir compte du nombre d'occurrences. L'échantillon terminologique est évalué en termes d'anglicité, de francité et de « lusité » ou « portuguité ». Si l'on compare le nombre d'entrées subissant « directement » l'influence de l'anglais par rapport au nombre d'entrées manifestant quelque souci de francisation ou de portuguisation de la part du locuteur, le panorama correspond au tableau 1 ci-dessous:

| | Français | | Portugais | |
|-------------------------|------------------|---------|-----------|--------|
| | Emprunts directs | 126/205 | 61,5 % | 95/146 |
| Emprunts par traduction | 79/205 | 38,5 % | 51/146 | 34,9 % |

Tableau 1: *Emprunts directs/emprunts par traduction*

À première vue, les résultats ne sont pas très probants. Il n'y a même pas 4 % de différence entre les deux langues en ce qui concerne la proportion d'emprunts

4 Le transfert d'image ([12] et [13]) relève d'une démarche beaucoup plus créative et les exemples sont rares. Il permet ici d'éviter un terme péjoratif, aussi bien en français qu'en portugais, pour traduire *sniff* [12] et de lever l'ambiguïté entre les états ou modes d'économie d'énergie *standby* [13] et *hold* (que le même locuteur traduit par le calque unilingue *estado espera*). Dans le premier cas (*standby*), l'appareil ne fait pas partie du réseau contrairement au second (*hold*), ce qui justifie le concept d'aliénation plutôt que celui de veille ou d'attente.

directs – selon un parcours purement sémasiologique – par rapport aux emprunts par traduction – selon le parcours sémasiologique du calque ou selon le parcours sémasio-onomasiologique du calque transpositionnel.

Cependant, si l'on compare le nombre d'emprunts intégraux, 100 % anglais, d'une part, au nombre de formes bilingues – qu'il s'agisse d'emprunt direct (*protocole L2CAP*), de calque (*couche L2CAP*) ou de calque transpositionnel (*pile protocolaire Bluetooth*) – et, d'autre part, au nombre de formes 100 % françaises ou portugaises, le tableau diffère légèrement :

| | Français | | Portugais | |
|------------------|----------|--------|-----------|---------|
| | 100 % EN | 56/205 | 27,3 % | 57/146 |
| Formes bilingues | 60/205 | 29,3 % | 44/146 | 30,14 % |
| 100 % FR/PT | 89/205 | 43,4 % | 45/146 | 30,82 % |

Tableau 2 : Degré d'internationalisation en français et en portugais

Ici, les emprunts intégraux l'emportent en portugais avec près de 40 % des entrées contre environ 30 % de formes bilingues et 30 % de formes 100 % portugaises, alors que c'est plutôt l'inverse en français, où plus de 40 % des entrées sont des formes 100 % françaises contre moins de 30 % de formes bilingues et encore moins d'emprunts intégraux.

La deuxième comparaison concerne la fréquence relative des procédés de formation secondaire, c'est-à-dire le nombre d'occurrences, dans chaque corpus (en français et en portugais ibérique), de chaque désignation française ou portugaise de huit concepts fondamentaux – et donc fréquents – de la technologie *Bluetooth* : les deux réseaux *ad hoc*, le statut de maître ou d'esclave au sein de ces réseaux et les quatre modes d'économie d'énergie. Ces huit concepts sont désignés en anglais par huit termes simples (*piconet, scatternet, master, slave, standby, hold, sniff, park*) ainsi qu'un dérivé (*parked*). Ces huit termes internationaux (y compris le dérivé) sont tous empruntés intégralement en français et en portugais, où ils servent notamment de références internationales.

Cette fois, la différence entre l'usage en français et en portugais du Portugal apparaît nettement (voir tableau 3 ci-dessous) : les francophones utilisent deux fois et demie plus de termes francisés que d'emprunts intégraux alors que les Portugais utilisent presque deux fois plus d'emprunts intégraux que de termes portugaisés.

| Termes EN | Français | | | | Portugais | | | |
|--------------------|------------------|--------|----------------|--------|------------------|--------|-----------------|--------|
| | Emprunt intégral | | Terme francisé | | Emprunt intégral | | Terme portugais | |
| <i>piconet</i> | 181 | 59 % | 126 | 41 % | 96 | 82,8 % | 20 | 17,2 % |
| <i>scatternet</i> | 33 | 82,5 % | 7 | 17,5 % | 19 | 76 % | 6 | 24 % |
| <i>master</i> | 20 | 5,4 % | 348 | 94,6 % | 52 | 45,2 % | 63 | 54,8 % |
| <i>slave</i> | 17 | 7,6 % | 207 | 92,4 % | 63 | 52,9 % | 56 | 47,1 % |
| <i>hold</i> | 15 | 53,6 % | 13 | 46,4 % | 10 | 90,9 % | 1 | 9,1 % |
| <i>park/parked</i> | 24 | 23,8 % | 77 | 76,2 % | 20 | 71,4 % | 8 | 28,6 % |
| <i>sniff</i> | 11 | 47,8 % | 12 | 52,2 % | 9 | 1000 % | 0 | 0 % |
| <i>standby</i> | 17 | 63 % | 10 | 37 % | 7 | 87,5 % | 1 | 12,5 % |
| Total | 318 | 28,4 % | 800 | 71,6 % | 276 | 64 % | 155 | 36 % |

Tableau 3 : Huit concepts en français et en portugais

En y regardant de plus près, on constate que l'usage de l'emprunt intégral ou d'un terme francisé ou portugais est relatif. Quatre emprunts intégraux sont plus fréquents en français, *piconet*, *scatternet*, *hold* et *standby*, alors qu'un seul terme portugais est plus fréquent en portugais, *mestre*.

En termes de concepts, *master* est le plus fréquent (483), il est suivi de *piconet* (423), *slave* (343), *park/parked* (129), *scatternet* (65), *hold* (39), *standby* (35) et *sniff* (32). Le poids du concept de *master* et de *piconet* tend ainsi à accentuer la différence puisque *mestre* contrarie la tendance générale en portugais et *piconet* en français.

Ces huit exemples montrent bien la complexité des facteurs qui se combinent pour influencer l'usage des locuteurs. Les mots courts et motivés comme *piconet* s'implantent naturellement à côté de *picoréseau* et *picorede* et les termes apparentés comme *scatternet* suivent le mouvement en dépit de leur opacité. Proportionnellement, *rede espalhada*, plus motivé, est plus fréquent en portugais (24 %) que les deux termes mal francisés, *réseau éclaté* ou *chaîné* (17,5 %), contrairement à la tendance générale. Les emprunts aménagés, courts et familiers, tels que *parqué* (76,2 %), ou techniques mais traditionnels, tels que *maître* (94,6 %) et *esclave* (92,4 %), dominent naturellement en français et l'usage est relativement plus équilibré dans ces trois cas en portugais. Les termes internationaux s'imposent en portugais et sont plus fréquents en français lorsque les formes naturalisées ou calquées sont péjoratives (la traduction officialisée par le groupe *Bluetooth renifleur* ne s'est pas implantée), plus longues (*scatternet* plutôt que le terme composé) ou ambiguës (*espera* équivaut aussi bien à *standby* qu'à *hold*, voir note 1; en français, l'ambiguïté se situe au niveau de *veille* et *attente*⁵). Toutefois, les efforts de désambiguïtation sont rares car il est plus simple d'adopter les termes internationaux univoques. On le voit,

5 Voir Bacquellaine (2008 : 109 et 111-112).

la motivation et la brièveté du néonyme sont effectivement des facteurs puissants d'implantation, mais ils ne sont pas les seuls.

Cette étude a donc confirmé l'influence de l'anglais sur les terminologies de pointe dans les langues supercentrales que sont le français et le portugais. Cette influence se manifeste non seulement par les emprunts directs ou par traduction, mais aussi par ce que Michael Picone (1996) appelle l'impératif synthétique. Dans le cas de la terminologie *Bluetooth* en français et en portugais, l'absence de joncteur est plutôt rare, mais les sigles sont extrêmement abondants, même s'ils s'appuient le plus souvent sur un noyau terminologique aménagé, redondant, comme dans protocole *L2CAP*, ou non, comme dans *transport ASB* (*Active Slave Broadcast*).

Certes, les termes anglais jouent indubitablement le rôle de référence internationale aussi bien en français qu'en portugais, mais on remarque toutefois que les locuteurs francophones recourent plus souvent que les Portugais à des mécanismes destinés à lever l'opacité des termes internationaux. Les efforts d'aménagement entrepris depuis plusieurs dizaines d'années et l'image que les locuteurs francophones se font de leur langue ont créé chez les francophones ce que l'on pourrait appeler un « réflexe de francisation » qui n'a pas encore d'équivalent chez les Portugais. Ce réflexe semble collectif chez les Québécois alors qu'il est individuel en Europe.

Si la terminologie *Bluetooth* est plus menacée en portugais ibérique qu'en français, les reformulations sont nécessaires dans ces deux langues. La terminologie *Bluetooth* existe donc aussi bien en français qu'en portugais. Le seul obstacle à sa diffusion est le besoin de la répertorier pour l'harmoniser, la mettre à la disposition du public et favoriser ainsi sa stabilisation dans les autres langues que l'anglais. Or, la terminographie fondée sur l'usage est une tâche complexe et infinie. Elle exige donc des ressources humaines et financières difficiles à mobiliser dans un monde où la langue est trop souvent considérée comme un outil secondaire et peu rentable.

2.4 Bilan

Alors, de deux choses l'une : soit les autres langues que l'anglais renoncent définitivement à se doter des moyens de traiter « tous les arts et toutes les sciences », un des objectifs de l'Académie française depuis sa fondation en 1635, soit le multilinguisme scientifique et technique se rétablit, ce qui n'est possible que grâce à un effort collectif, réalisé non seulement par l'État, mais surtout par le citoyen, acteur principal de l'évolution et de la survie des langues. Les résultats des efforts d'aménagement terminologique, réalisés notamment par l'État français depuis une quarantaine d'années, révèlent la difficulté de l'intervention de l'État sur la langue et entraînent une redéfinition du rôle de l'État et du rôle du citoyen.

L'État doit promouvoir l'image que les locuteurs se font de leur langue. Il doit aussi permettre au plus grand nombre l'accès au savoir en assurant un enseignement efficace, fondé sur une terminologie accessible. Il peut également participer à la normalisation terminologique au sein d'organismes de normalisation nationaux

ou internationaux. Il peut encore subventionner la publication scientifique dans la ou les langue(s) nationale(s), le tout est de convaincre les scientifiques d'écrire dans une autre langue que l'anglais. Il peut aussi légiférer pour que les documents techniques accompagnant les produits commercialisés sur son territoire soient accessibles aux consommateurs dans leur langue. Il peut enfin sensibiliser les producteurs de technologies de pointe à la plus-value terminologique, mais il ne peut se charger de combler le déficit terminologique dans les domaines de pointe et les organismes de normalisation ne le peuvent pas non plus.

Les producteurs de technologies de pointe sont eux-mêmes des citoyens qui devraient songer à confier la tâche terminographique de leur domaine à des collègues d'experts et de linguistes chargés de fournir à leurs concitoyens les outils nécessaires au discours scientifique et technique dans d'autres langues que l'anglais. Ainsi, le citoyen convaincu de la valeur de sa langue et de l'importance du maintien de toutes ses fonctions pourrait l'utiliser pour traiter « tous les arts et toutes les sciences ».

Conclusion

Les résultats de cette étude portent à croire que les langues romanes sont moins menacées par la perte de domaines que les langues germaniques, apparentées – et donc plus perméables – à l'anglais. Certes, les termes anglais servent de référence en français comme en portugais, mais ils sont plus opaques que dans les langues germaniques et requièrent presque systématiquement des reformulations, notamment dans le discours didactique. Les procédés de formation secondaire en français et en portugais sont similaires, mais les habitudes langagières diffèrent sensiblement d'une communauté linguistique à l'autre. Le rayonnement terni du français dans le monde lui réserve encore une place de choix sur le « marché aux langues » de Calvet, ce qui lui confère un plus grand prestige auprès de ses locuteurs natifs que celui dont jouit le portugais au Portugal. Il en résulte un réflexe de francité sans véritable équivalent en portugais, mais cela peut changer. L'accord sur l'orthographe entre le Brésil et le Portugal est entré en vigueur le 1^{er} janvier 2009, ce qui élargit la communauté et l'espace lusophone réunifiés et hausse le prestige du portugais auprès de ses locuteurs.

La préservation de la diversité linguistique doit être globale : elle passe par les Xphonies, par l'État, mais surtout par le citoyen : seules survivent et se développent les langues qui se parlent. Les terminologies de pointe sont des outils indispensables au discours scientifique et technique. Sans terminologie propre, le discours scientifique ou technique dans une autre langue que l'anglais devient impossible et la langue perd peu à peu des domaines, comme dans les langues germaniques, ou alors le discours se transforme en un jargon hybride imprononçable, comme en français et en portugais, à moins que le locuteur – le citoyen – ne fasse preuve de créativité pour rendre son discours plus naturel. Ainsi naît la « néonymie sauvage » qui pourrait être « apprivoisée » grâce à un effort terminographique global,

en divisant cette tâche colossale entre les producteurs de technologies de pointe. Cet effort contribuerait certainement à la préservation de la diversité linguistique.

Bibliographie

- Bacquelaine (Fr.), 2008 : *La terminologie Bluetooth en anglais, en français et en portugais. Étude de néonymie comparée*, Porto, 30 septembre 2008, aleph. letr. up. pt/exlibris/aleph/a18_1/apache_media/F82XENV SXBAPYAT6NHAIKKCJQJ88P. pdf
- Calvet (L.- J.), 1999 : *Pour une écologie des langues du monde*, Paris, Plon.
- Calvet (L.- J.), 2000 : « Vie et mort des langues : les locuteurs décident » dans *Le Courrier de l'Unesco, Dossier La paix des langues*, avril 2000, p. 35-36, www.unesco.org/courier/2000_04/fr/doss25.htm
- Calvet (L.- J.), 2002 : *Le marché aux langues. Les effets linguistiques de la mondialisation*, Paris, Plon.
- Chansou (M.), 2003 : *L'aménagement lexical en France pendant la période contemporaine (1950-1994), Étude de sociolexicologie*, Paris, Champion.
- Conceição (M.C.), 2005 : *Concepts Termes et Reformulations*, Lyon, Presses universitaires de Lyon.
- Depecker (L.), 2001 : *L'invention de la langue. Le choix des mots nouveaux*, Paris, Armand Colin.
- Depecker (L.), 2005 : « Contribution de la terminologie à la linguistique », dans *Langages* 157, La terminologie : nature et enjeux, Depecker (L.), dir., mars 2005, Paris, Larousse, p. 6-13.
- Depecker (L.) et Dubois (V.), éd., 2005 : *Les néologies contemporaines*, Actes du colloque tenu le 15 octobre 2004 à L'École normale supérieure de la rue d'Ulm, Paris, Société française de terminologie (Le savoir des mots).
- Faraco (C.A.), org., 2002 : *Estrangeirismos : guerras em torno da língua*, 2^e éd., São Paulo, Parábola.
- Glissant (E.), 1997 : *Traité du Tout-Monde*, Paris, Gallimard.
- Klinkenberg (J.-M.), 2001 : *La langue et le citoyen. Pour une autre politique de la langue française*, Paris, PUF.
- Linguatca, Pólo FLUP/CLUP, 2008 : *Corpógrafo V4*, Porto, www.linguatca.pt/corpografo/.
- Picone (M.), 1996 : *Anglicisms, Neologisms and Dynamic French*, Amsterdam, John Benjamins Publishing.
- Walter (H.), 1997 : *L'aventure des mots français venus d'ailleurs*, Paris, Robert Laffont (Le Livre de Poche).
- Walter (H.), 2001 : *Honni soit qui mal y pense. L'incroyable histoire d'amour entre le français et l'anglais*, Paris, Robert Laffont.

Taoufik Massoussi,
LDI (UMR 7187), viavoo SAS,
Villetaneuse, Boulogne-Billancourt,
France,
taoufik.massoussi@viavoo.com.

Soumaya Mejri,
LDI (UMR 7187),
Université de Jendouba,
Tunis,
Tunisie,
soumayamejri@yahoo.fr.

Traitement automatique des discours financiers : application à la bourse

Résumé

Dans cet article, nous exposons les éléments méthodologiques qui président à l'étiquetage automatique des textes de la bourse. Ces éléments ont trait à la constitution de ressources linguistiques et à leur implémentation dans des systèmes d'extraction d'information textuelle. Les différentes étapes du traitement sont présentées. Elles montrent les solutions adoptées face aux différentes difficultés, et donnent lieu à des perspectives théoriques et pratiques dans le domaine du traitement automatique des langues naturelles (TALN).

Mots-clés : bourse ; étiquetage sémantique ; dictionnaire électronique ; grammaire locale ; prédicat ; inférence.

Introduction

Ce travail s'inscrit dans un cadre plus général qui consiste à traiter les discours stratégiques de façon générale, et à étudier l'impact qu'un discours peut avoir sur un autre. Une recherche est actuellement en cours sur l'impact du discours des entreprises sur le discours des analystes financiers (Mejri à paraître). Dans cet article, il s'agit d'étudier les bulletins boursiers publiés dans les journaux spécialisés en finance, et qui rendent compte de l'état de santé de la bourse – et plus généralement des marchés financiers – et décrivent le mouvement des indices boursiers. Notre objectif est de développer un système d'étiquetage automatique des textes financiers, dont la performance sera doublement évaluée : (i) par le taux de résolution des ambiguïtés morfo-syntaxiques, et sémantiques et (ii) par son adéquation

aux besoins des utilisateurs de ce système. C'est pour cela qu'un module de traduction des étiquettes métalinguistiques en étiquettes informationnelles, adapté à ces besoins, est prévu, *cf. infra*.

Sur plan théorique, l'objectif est de montrer qu'une analyse sémantique automatique ne peut se faire ni indépendamment de la linguistique, ni indépendamment d'une prise en compte des faits de langue rencontrés. Nous souhaitons établir l'idée que les systèmes développés en ingénierie linguistique sont plus performants quand ils intègrent des outils fondés sur des ressources linguistiques fines, pour peu que la constitution de ces ressources soit effectuée dans un cadre méthodologique cohérent, adapté aux exigences épistémologiques des systèmes de traitement de l'information textuelle. La méthode adoptée est donc strictement linguistique. Elle consiste à élaborer des ressources linguistiques complètes et à les projeter sur les textes du corpus. Nous reviendrons sur les principaux outils théoriques de ce modèle, nécessaires à la compréhension de cet article. Pour le reste, nous renvoyons aux différents travaux qui se réclament de ce modèle (*cf. bibliographie*).

Après avoir rappelé brièvement l'état des lieux en matière de traitement automatique des discours spécialisés, nous caractériserons notre corpus et la méthode de collecte adoptée. Ensuite, nous présenterons les outils utilisés pour l'étiquetage. Nous détaillerons les différentes étapes de l'analyse à l'aide de transducteurs¹ (Silberztein 1993). Enfin, nous aborderons la problématique de l'inférence dans les discours financiers.

1 État des lieux en terminologie : corpus et dictionnaires

Deux approches sont privilégiées dans le traitement automatique des langues de spécialité : une approche statistique, utilisée pour le traitement des grands corpus, et une approche ontologique utilisée le plus souvent pour le traitement de corpus multilingues.

1.1 Approches statistiques pour le traitement des corpus de spécialité

Le postulat de la démarche statistique consiste à corrélérer la signification des unités lexicales à leur fréquence d'apparition. Plusieurs études en statistiques lexicales font l'hypothèse que les informations nécessaires à la reconnaissance du sens d'un mot sont définies par ses environnements stricts, dégagés par des logiciels d'extraction opérant sur de vastes corpus. Or, cette recherche de cooccurrences ne peut pas être mécanique. Un programme qui extrait les caractères qui se trouvent dans le voisinage immédiat d'un verbe, par exemple, ne saurait constituer des conditions nécessaires à l'interprétation correcte de ce verbe. En effet, les éléments contextuels qui définissent un mot ne lui sont pas nécessairement contigus. Dans les textes, des insertions de diverses natures peuvent se trouver entre un sujet et son verbe :

¹ Les transducteurs sont des automates à états finis qui permettent de reconnaître et d'étiqueter par des informations métalinguistiques, des unités lexicales correspondant à des configurations spécifiques. Nous utilisons la plate-forme Unitex, dans sa version 2.0, élaborée par Max Silberztein.

L'entreprise, nous le pensons fort, doit poursuivre sa stratégie de croissance rentable.

Le travail d'analyse doit donc commencer par traiter ces contextes aléatoires, comme les incises (Fairon 2000). Une deuxième difficulté réside dans la reconnaissance des constructions figées qui, d'un point de vue purement statistique, constituent de faux environnements, puisque le sens du verbe dans des locutions verbales, par exemple, n'est pas déterminé par son complément (*prendre une veste, casser sa pipe, jeter l'éponge*). C'est plutôt l'ensemble qui forme une unité de sens.

1.2 Pratique des dictionnaires de terminologie actuels

Le traitement automatique des langues de spécialité ne peut pas se fonder sur la pratique actuellement en usage en terminologie, qui ne répond pas aux spécificités épistémologiques des systèmes informatiques. Si on examine les dictionnaires de spécialité, on trouve majoritairement des noms, le plus souvent composés, très peu d'adjectifs et quasiment pas de verbes ou d'adverbes. Qu'ils soient monolingues, bilingues ou multilingues, ces dictionnaires présentent un déséquilibre entre les termes strictement spécialisés et les termes issus de la langue générale. Ce qui sous-tend ce déséquilibre, c'est une certaine idée courante en terminologie qui consiste à voir dans les langues spécialisées des vocabulaires complètement isolés de la langue générale. Or, dans les corpus, on observe une présence massive des termes de la langue générale. La majorité de ces termes sont polysémiques et, pour reconnaître chaque signification, on doit être en mesure de décrire chaque emploi avec beaucoup de précision. Aucun traitement automatique n'est envisageable si on n'a pas une description fine de la langue générale.

S'agissant de notre corpus, aucune des deux méthodes n'est satisfaisante pour au moins deux raisons: (i) les textes financiers sont des textes courts, et les éléments significatifs apparaissent souvent une seule fois, (ii) une grande partie du vocabulaire de la bourse est constitué d'unités lexicales issues de la langue générale. Le taux de silence dans l'étiquetage basé sur la fréquence ou des ontologies de domaine est donc beaucoup trop important.

2 Présentation du corpus

2.1 La bourse

La bourse est définie comme un lieu de financement et de placement. Elle remplit plusieurs fonctions:

- le financement des entreprises: le marché boursier permet aux entreprises de se financer, d'investir, en mettant directement en contact l'offre et la demande de capital. On parle de système désintermédié à la différence du système bancaire. Par l'introduction en bourse, les entreprises ont accès à des ressources financières supplémentaires;
- un lieu d'épargne: les investisseurs ayant une capacité d'épargne positive peuvent devenir les actionnaires des entreprises privées ou devenir les créanciers de

ces sociétés et des collectivités publiques. Ils peuvent investir soit à l'occasion d'augmentations de capital ou d'introductions en bourse sur le marché primaire, ou s'orienter sur le marché dit secondaire sur lequel sont négociés les titres déjà émis. La bourse permet la mobilisation de l'épargne et, partant, une meilleure allocation des ressources vers les projets d'investissement les plus profitables. En facilitant l'accès à l'épargne, les bourses rendent possibles de nouveaux investissements et améliorent donc la rentabilité des investissements.

- la gestion du risque: la bourse permet de transférer le risque par le biais des négociations d'actions et d'obligations. Elle permet également aux investisseurs de se protéger du risque par l'utilisation de produits dérivés: les *swaps*, les *forwards*, les contrats à terme, les options, etc. ;
- la liquidité des titres: la bourse permet la négociabilité des titres. Cette facilitation de la sortie permet d'attirer un plus grand nombre d'investisseurs et explique pourquoi les sociétés de capital-risque souhaitent que les entreprises dans lesquelles elles ont investi s'introduisent en bourse ;
- et la mesure économique: les cotations effectuées à la bourse permettent de mesurer la valeur attribuée par le marché à une entreprise, à une matière première ou à une créance dans le temps. Les cours permettent donc de suivre l'évolution du prix de produits et l'évolution générale de la situation économique d'un pays.

Ce positionnement de la bourse dans le monde de la finance est corrélé à des caractéristiques linguistiques auxquelles un système d'analyse automatique doit s'adapter.

2.2 Caractérisation linguistique du corpus

Les textes financiers relatifs au marché boursier décrivent essentiellement :

- le mouvement des indices boursiers :

[1] Paris poursuit sa hausse dans le sillage de Wall Street.

[2] Le dollar chute brutalement.

- et l'état psychologique des bourses et des marchés financiers :

[3] La Bourse de Paris s'est fait peur.

L'actualisation de ces prédicats est souvent prise en charge par des verbes métaphoriques :

[4] Les fluctuations du dollar ravivent les inquiétudes des marchés.

Concernant les arguments, on note un taux de polysémie très élevé. Pour parler du mouvement ou des sentiments, le discours sur la bourse emploie indifféremment des noms de lieux, des noms de monnaies, des noms humains, etc. Si on prend

l'exemple du nom *Wall Street*, on se rend compte qu'un grand nombre de classes d'arguments employées dans ces textes ont plusieurs emplois métonymiques :

- [5] Wall Street est fermé aujourd'hui pour cause de jour férié aux États-Unis : <établissement>
- [6] Wall Street a clôturé à la baisse aujourd'hui : <bourse>
- [7] Wall Street veut croire à la fin de la crise : <personnel de direction>
- [8] Wall Street est en grève depuis lundi : <employés>
- [9] Wall Street a franchi le plafond des 10 000 points : <indice boursier>

L'apport du système d'analyse automatique de ce genre de textes ne se situe donc ni dans la réduction de la longueur des textes, ni dans la génération de résumés, mais plutôt dans la qualité des analyses sémantiques, et de façon transversale, dans la prise en compte des phénomènes de la métaphore et de la métonymie. Il faut donc disposer d'une méthodologie linguistique capable de construire des ressources linguistiques générales qui prennent en compte, dans la définition des prédicats, la nature de leurs arguments et de leurs actualisateurs.

3 Dictionnaires électroniques

Nous rappelons les principales caractéristiques méthodologiques des dictionnaires que nous utilisons pour l'étiquetage des textes de la bourse. Elles ont trait essentiellement à deux niveaux : le niveau structural qui définit la relation entre les prédicats et les arguments, et le niveau de l'actualisation.

3.1 L'unité minimale d'analyse

Le modèle des classes d'objets postule qu'une phrase élémentaire est l'unité minimale d'analyse d'une langue spécialisée. Toute phrase élémentaire est constituée d'un prédicat de premier ordre et de ses arguments. Dans la phrase suivante :

- [10] Renault vend des voitures à l'étranger.

le verbe *vendre* est un prédicat de premier ordre. Il sélectionne trois arguments, respectivement : *Renault*, *voitures* et *étranger*. Cette phrase donne lieu à la représentation suivante :

VENDRE (Renault, voitures, étranger)

Un texte ne peut évidemment pas être réduit à une suite de phrases élémentaires. Il comprend des connecteurs, des réductions de schémas d'arguments, des reprises anaphoriques, etc. Il reste que le recensement et la description des phrases simples sont les premières des conditions d'un traitement automatique.

3.2 La notion d'emploi

Cette notion est pratiquée de façon plus ou moins systématique par les dictionnaires. Pour déterminer de façon précise le sens d'un mot et en donner une définition, il faut indiquer son emploi. L'emploi d'un prédicat, c'est l'ensemble des noms qu'il sélectionne en position sujet et/ou complément. Un même prédicat morphologique aura autant d'emplois qu'il a de types sémantiques d'arguments. Inversement, l'emploi d'un argument est déterminé par la liste des prédicats (verbes, adjectifs, noms) qui le sélectionnent. Bien souvent, pour catégoriser les arguments et distinguer les significations d'une même forme prédicative, les dictionnaires se contentent de la notion de traits de sélection (*humain, concret, locatif*, etc.). Bien que cette information soit capitale dans l'opération de désambiguïsation, elle reste trop puissante, car elle permet de reconnaître et de générer des phrases mal formées. Le même principe préside à la description des prédicats. Ces derniers sont réunis dans des classes sémantiques.

3.3 La notion de « classe » : classes d'arguments et classes de prédicats

La notion de classe a l'avantage de regrouper un grand nombre de phrases dans le cadre d'un emploi déterminé et de leur proposer un traitement uniforme en mettant en évidence toutes les propriétés communes. C'est ainsi que nous évitons de répéter la même instruction pour les emplois similaires, et que nous l'indiquons à la classe entière. Ce compactage de l'information concerne tant les arguments que les prédicats. C'est ce niveau de description qui est le plus pertinent dans un dictionnaire électronique. Ces classes sémantiques ne sont pas de nature ontologique, mais proviennent d'analyses linguistiques. Elles sont fondées sur les descriptions préalables des différents emplois. Par exemple, tous les prédicats sémantiquement contigus et qui ont des propriétés syntaxiques communes sont réunis dans des classes sémantiques. Le principe d'homogénéité est fondamental dans l'établissement des classes sémantiques. À partir de sous-classes homogènes décrites en intension (l'ensemble des propriétés) et en extension (la liste des éléments de la classe), il est possible de faire des regroupements à des niveaux supérieurs, celui des classes et celui des hyperclasses.

3.4 L'actualisation

L'actualisation désigne le procédé linguistique qui permet non seulement d'inscrire la phrase constituée d'un prédicat et de ses arguments dans les catégories grammaticales telles que le temps, l'aspect, la modalité, le genre, le nombre, la voix, etc., mais aussi de contribuer à son interprétation (Gross et Vivès 1986, Buvet à paraître). Il s'agit donc d'un paramètre fondamental dans le dispositif théorique adopté ici, et dans le traitement des discours financiers, qui constitue notre application.

Parmi les trois dictionnaires² élaborés suivant cette méthodologie, le dictionnaire des actualisateurs se distingue par le fait que ses entrées correspondent à des emplois regroupés non pas dans des classes sémantiques, mais dans des classes

2 Dictionnaire des prédicats, des arguments et des actualisateurs.

grammaticales. De ce point de vue, la distinction entre signification lexicale et signification grammaticale est essentielle. Par exemple, s'agissant de l'actualisation nominale, il s'agit de décrire tous les verbes supports appropriés à des classes de prédicats :

- Prédicats d'<affects> : *exploser de joie, se consumer de haine, éprouver de la peur*
- Prédicats d'<aides financières> : *accorder un crédit, arroser <qqn> de subventions*

Dans la même classe de prédicats, on décrit aussi bien des actualisateurs standards (*éprouver de la joie*) et des actualisateurs non standards (*exploser de joie*). Cette distinction qui ressort des travaux sur l'actualisation (Buvet à paraître) permet de montrer que les verbes métaphoriques, lorsqu'ils jouent le rôle d'actualisateurs, véhiculent toujours une signification supplémentaire.

Ces éléments méthodologiques ont un impact direct sur la reconnaissance automatique des phrases de la bourse par des automates. Après avoir identifié des prédicats et des arguments, la reconnaissance des phrases consiste à spécifier la nature de tous les autres éléments qui ont trait à l'actualisation. Il s'agit alors de les exclure en tant que tels et de leur associer, s'il y a lieu, un code relatif à leur fonction dans la phrase. Les deux phrases suivantes :

[11] Paris est en léger repli.

[12] Paris accentue ses pertes.

est représentée comme suit :

ATTENUATIF en repli (Paris)

INTENSIF pertes (Paris)

4 Phases du traitement

Les différentes étapes du traitement consistent à étiqueter sémantiquement des éléments du texte et à identifier des tournures à l'aide des outils de la plateforme Unitex. Les ressources sont des dictionnaires électroniques et des transducteurs qui peuvent ou non faire appel à ces dictionnaires. Les arguments font l'objet d'une description lexicographique tandis que les prédicats sont décrits sous forme de grammaires locales.

Ces étapes ont trait à la gestion du bruit (détecter un <affect> ou un <mouvement > alors qu'il n'y en a pas) et du silence (ne pas reconnaître un <affect> ou un <mouvement> alors qu'il y en a un).

4.1 Prétraitement

La première étape consiste à appliquer des transducteurs dont le rôle est de traiter les entités nommées³ et certains adverbess figés. Certains noms de personnes sont, en effet, ambigus et peuvent être confondus avec des <affects> : *Jean-Yves Ravis*, *Claudette Sérénité*, *Zoé Félicité*. La solution est alors de construire un graphe qui fait appel au dictionnaire des prénoms et de l'associer aux noms d'<affects> et d'étiqueter l'ensemble comme une entité nommée.

D'autre part, certains adverbess complexes, du type *en dépit de* peuvent être confondus avec des <affects> : *La montée du dollar cause du dépit à certains investisseurs*.

Nous construisons à cet effet des graphes de reconnaissance des séquences figées. Partant de la double structuration de ces séquences, c.-à-d. une structure interne polylexicale et une structure externe qui correspond à un fonctionnement monolexical, il s'agit d'éviter à l'automate de reconnaître comme des affects ou des mouvements des suites dont la structure interne contient un mot de sentiment ou de mouvement (*en dépit de*, *enquête de satisfaction*, etc.) et de reconnaître comme des suites figées des suites du type (*avoir le cafard*, *donner froid dans le dos*, *perdre les pédales*, etc.). Comme on l'a dit ci-dessus, ces transducteurs permettent de gérer respectivement le bruit et le silence. La figure suivante illustre le résultat des transducteurs qui reconnaissent les suites du premier type.

| RÉSULTAT PRETRAITEMENT DES SÉQUENCES FIGÉES | | |
|---|--|---------------------------------------|
| Cette chute s'explique | <u>bien sûr [LOCUTION OUT]</u> | par le ralentissement économique |
| des candidats sont | <u>bien sûr [LOCUTION OUT]</u> | examinées lors du vote de l'assemblée |
| sous l'égide de la Société générale, | <u>en dépit de [LOCUTION OUT]</u> | son taux nominal assez bas (9%) |
| économiques américains.(5) Le Japon, | <u>en dépit de [LOCUTION OUT]</u> | difficultés financières à court terme |
| Ce n'est | <u>pas la peine de [LOCUTION OUT]</u> | répéter aujourd'hui l'erreur du passé |
| l'ouverture du marché du riz | <u>sous peine de [LOCUTION OUT]</u> | faire capoter l'Uruguay Round. |
| du CAC 40 vont lancer des | <u>enquêtes de satisfaction [LOCUTION OUT]</u> | auprès de leurs clients, afin de |

Figure 1

4.2 Reconnaissance des arguments

La deuxième étape consiste à reconnaître les arguments. Les transducteurs qui étiquettent les arguments font appel à des dictionnaires. Plusieurs dictionnaires sont utilisés : les noms de <pays>, les noms de <marques>, les noms d'<organismes financiers>. Ces dictionnaires sont appliqués à chaque fois que l'on charge un bulletin boursier. Un travail de description des prédicats métonymiques permet de reconnaître le contexte où ces arguments sont interprétés en tant que collectifs humains (Massoussi 2008a) :

3 Noms de personnes, d'organisations, d'entreprises, de lieux, de dates, etc.

| RÉSULTAT TRAITEMENT DES ARGUMENTS MÉTONYMIQUES | |
|--|--|
| [BANQUE_RESPONSABLE_FINANCIER] | <u>La BCE</u> se refuse d'aller jusqu'à un taux zéro. |
| [BANQUE_RESPONSABLE_FINANCIER] | <u>La BNP</u> a peu de chances de convaincre les actionnaires de Fortis. |
| [BANQUE_RESPONSABLE_FINANCIER] | <u>La Caisse d'épargne</u> reporte la suppression de 4.500 postes. |
| [BANQUE_RESPONSABLE_FINANCIER] | <u>HSBC</u> va racheter les minoritaires de sa filiale costaricaine pour 7 millions. |
| [CAPITALE_RESPONSABLE_POLITIQUE] | <u>Londres</u> va s'intéresser de plus près au système des bonus. |
| [CAPITALE_RESPONSABLE_POLITIQUE] | <u>Paris</u> et Berlin prônent un rapport d'étape rapide dans l'Union |
| [CAPITALE_RESPONSABLE_POLITIQUE] | <u>Pékin</u> croit pouvoir atteindre les 8% de croissance en 2009. |
| [MARQUE_GROUPE INDUSTRIEL] | <u>Coca-Cola</u> investit 2 milliards de plus en Chine sur 3 ans |
| [CAPITALE_RESPONSABLE_POLITIQUE] | <u>Ford</u> réduit sa dette via des rachats d'obligations |
| [CAPITALE_RESPONSABLE_POLITIQUE] | <u>General Motors</u> n'exclut pas le scénario d'une liquidation |
| [CAPITALE_RESPONSABLE_POLITIQUE] | <u>Nissan</u> suspend sa participation dans l'usine de Tanger |
| [CAPITALE_RESPONSABLE_POLITIQUE] | <u>Porsche</u> reconnaît avoir emprunté 700 millions d'euros |
| [CAPITALE_RESPONSABLE_POLITIQUE] | <u>Volvo</u> discute de l'octroi d'une garantie avec la région flamande |
| [PAYS_RESPONSABLE_POLITIQUE] | <u>L'Allemagne</u> hésite à aider les grands groupes industriels |
| [PAYS_RESPONSABLE_POLITIQUE] | <u>La Corée du Nord</u> annonce un nouvel essai nucléaire |
| [PAYS_RESPONSABLE_POLITIQUE] | <u>Les Etats-Unis</u> veulent « écouter » le reste du monde |
| [PAYS_RESPONSABLE_POLITIQUE] | <u>Les Etats-Unis</u> proposent un match amical à l'Iran |

Figure 2

4.3 Reconnaissance des prédicats

Lors de cette étape, on gère les ambiguïtés sémantiques (polysémie), en étiquetant selon la configuration syntaxique (environnement gauche-droite), les formes susceptibles d'avoir des emplois en tant que prédicats d'<affects> ou de <mouvements>. Grâce aux transducteurs, on reconnaît les prédicats et on les étiquette en fonction de la classe sémantique à laquelle ils appartiennent.

| RÉSULTAT TRAITEMENT DES PREDICATS D'AFFECTS | | |
|---|--|-------------------------------------|
| Les analystes de BAM Lynch | <u>craignent</u> [AFFECT_CRAINTE] | que le secteur d'activités de |
| Le moral reste inchangé en raison des | <u>craintes</u> [AFFECT_CRAINTE] | liées à la montée du chômage |
| Cette remontée des cours apaise les | <u>craintes</u> [AFFECT_CRAINTE] | sur la demande |
| Affaiblies hier par les | <u>craintes</u> [AFFECT_CRAINTE] | concernant leurs dettes |
| La baisse des commandes industrielles | <u>a déçu</u> [AFFECT_DECEPTION] | pour le mois de mars. |
| La seconde partie de séance est | <u>calme</u> [AFFECT_TRANQUILLITE] | à la Bourse de Paris. |
| Dans la zone euro, la baisse des commandes | <u>a déçu</u> [CAUSATIF_AFFECT_DECEPTION] | pour le mois de mars. |
| Ceux-ci subissaient des nouvelles | <u>inquiétantes</u> [CAUSATIF_AFFECT_INQUIETUDE] | en provenance de la Corée du |
| Les investisseurs | <u>s'inquiètent</u> [AFFECT_INQUIETUDE] | de l'annonce d'une augmentation de |
| Cette remontée des cours | <u>apaise</u> [CAUSATIF_AFFECT_APAISEMENT] | les craintes sur l'orientation de |
| La Bourse de New-York, | <u>ravie</u> [AFFECT_PLAISIR] | de la progression de l'indice de |
| La | <u>déprime</u> [AFFECT_CAFARD] | des ménages se propage aux marchés |
| Le | <u>soulagement</u> [AFFECT_APAISEMENT] | sur le marché tokyoite était amplif |
| La tendance : Un marché qui oscille entre | <u>euphorie</u> [AFFECT_EUPHORIE] | et désespoir |
| Un marché qui oscille entre euphorie et | <u>désespoir</u> [AFFECT_DESEPOIR] | |
| Dans ce contexte, la météo boursière | <u>maussade</u> [AFFECT_MAUSSADERIE] | de ce début de saison estivale, où |
| Le cuivre, le métal rouge qui | <u>affole</u> [CAUSATIF_AFFECT_TROUBLE] | la bourse |

Figure 3

4.4 Traitement structurel

Certaines structures, principalement la négation, changent le sens des prédicats. Lors de cette étape, on représente donc les différentes configurations morpho-syntaxiques de la négation et on l'applique aux prédicats reconnus pour en changer la

valeur. D'autres marqueurs comme les semi-auxiliaires (*éviter, empêcher*) ou certains adverbes (*loin*) exprimant une négation sémantique sont aussi représentés :

[13] L'heure n'est pas à l'euphorie dans la Bourse de Wall Street.

[14] On est loin de l'agitation des derniers mois.

[15] Wall Street évite la panique malgré les chiffres de l'emploi.

4.5 Du sens à l'information

Actuellement, nous travaillons sur un système de correspondance entre des étiquettes sémantiques (classes de prédicats et classes d'arguments) et des étiquettes informationnelles. Si les étiquettes sémantiques consistent à indiquer pour chaque prédicat reconnu sa classe sémantique, les étiquettes informationnelles sont de deux sortes : pour les prédicats d'<affects>, il s'agit d'indiquer si c'est un affect positif ou négatif, tandis que pour les <mouvements>, on indique la direction (HAUT, BAS) et la vitesse (RAPIDE, LENT). Ces étiquettes devraient fournir des informations plus exploitables par les systèmes d'analyse financière.

L'intérêt porté à l'analyse automatique des discours financiers se fait dans le but d'un meilleur déchiffrement et d'un meilleur décryptage de la bonne information financière. Dans le contexte actuel de crise financière et économique où règne un climat de méfiance, la bonne information se fait de plus en plus rare. Elle devient précieuse pour les investisseurs et sa détection constitue la priorité de tous les acteurs financiers. Comme les financiers ont tendance à ne s'intéresser qu'aux chiffres et à les présenter en vrac, ils se montrent très négligents et non vigilants dans leur façon de s'exprimer. Ce dernier point, qui demeure encore inexploité en sciences de gestion, présente un créneau très intéressant.

5 Perspectives : traitement des inférences

L'examen d'un grand nombre de textes montre que l'inférence est un paramètre très structurant. Si on définit l'inférence comme le mécanisme logique par lequel on rétablit une prédication B (sous-jacente), à partir d'une prédication A (réalisée en surface), on peut voir qu'elle a un lien étroit avec la notion de codage (Gross et Prandi 2004). Deux types d'inférences causales peuvent être traités dans notre corpus. Ces deux types ont trait à la relation de cause, qui s'exprime dans des phrases complexes : il s'agit soit d'un enrichissement inférentiel, soit de cas de parataxe.

Nous renvoyons aux travaux de Gaston Gross (2004, à paraître) sur les relations transphrastiques pour :

- la remise en cause de l'analyse classique de la phrase complexe en termes de proposition ;
- la critique de l'analyse des connecteurs comme des séquences figées ;
- le rôle que jouent les substantifs et les adjectifs dans ces connecteurs ;

- et la hiérarchie prédicative entre prédication de second ordre et prédication élémentaire.

Une avancée théorique majeure résulte de ces travaux. Elle consiste à montrer qu'il y a des restrictions de sélection entre les prédicats de second ordre et les prédicats élémentaires. Ces restrictions ont trait soit à la polarité (négative, positive) ou aux classes sémantiques. La projection des prédicats de cause sur un grand corpus (dix années du journal *Le Monde*) montre que le spectre nominal d'un verbe comme *causer* est constitué essentiellement d'événements négatifs (*mort* [602]; *dégât* [223]; *perte* [123]; *dommage* [97]; *préjudice* [88]), tandis que *susciter* se combine préférentiellement avec des actions ou des états humains (*réaction* [864]; *inquiétude* [551]; *intérêt* [438]; *interrogation* [396]; *débat* [368]). Le traitement de l'inférence dans les bulletins boursiers repose sur les résultats de ces travaux. D'abord, la typologie établie des prédicats de cause nous permet d'élaborer des grammaires locales pour le traitement de l'enrichissement inférentiel. Ensuite, nous avons observé une relation d'appropriation entre les affects et les mouvements. Une explicitation des inférences causales, en cas de parataxe, devient alors possible.

5.1 Enrichissement inférentiel

Pour expliquer tel mouvement d'un indice boursier ou tel état psychologique des marchés financiers, plusieurs bulletins boursiers établissent un lien temporel entre une affirmation ou une déclaration faite par un responsable politique ou économique et un mouvement ou un état psychologique de la bourse :

- [16] La *Bourse* de Paris a brutalement chuté mercredi suite aux rumeurs qui ont circulé sur des dépréciations d'actifs à la Société générale en France.
- [17] La place parisienne a immédiatement pris la voie de la baisse après le discours de Ben Bernanke, qui confirme la dégradation de la conjoncture économique américaine.

Grâce à la description des prédicats de second ordre du type *suite à*, *être lié à*, etc., il est possible d'étiqueter ces phrases comme étant reliées par des relations de cause. La procédure en question est un enrichissement inférentiel (König et Traugott 1988; Hopper et Traugott 1993).

Un programme de travail consiste alors à récupérer le contexte gauche et droit de tous les relateurs de temps (*suite à*, *du moment que*, *après*), de fréquence (*à chaque fois que*, *toutes les fois que*), d'hypothèse (*si p*, *alors q*) et de faire la typologie des prédicats figurant dans ces contextes. Cette classification permettra de corrélérer statistiquement tel discours ou telle décision à un mouvement ou à un affect donné.

5.2 Inférences et parataxe

Si nous prenons la phrase suivante :

[18] La Bourse de Paris s'est fait peur: le CAC 40 est passé brièvement sous les 2500 points,

on observe une relation de cause entièrement inférée, dans la mesure où elle n'est prise en charge par aucun élément lexical. C'est la connaissance spécialisée du domaine de la bourse qui nous dit que l'état psychologique de la bourse dépend étroitement des mouvements des indices et des monnaies. Le seul moyen de détecter automatiquement ces inférences et de disposer d'une description complète des deux classes de prédicats.

6 Applications envisagées

Les applications que l'on peut envisager à partir des ressources linguistiques constituées en rapport avec les textes financiers sont toutes fondées sur une phase d'interprétation des documents. Celle-ci implique une analyse linguistique des rapports financiers qui donne lieu à une représentation interprétée par la suite selon les finalités de l'application voulue. Parmi ces applications, on peut citer :

- la veille stratégique: il s'agit de chercher des informations ciblées dans les textes. Le système de veille fonctionne de façon continue – contrairement aux moteurs de recherche qui fonctionnent de façon discontinue – et permet d'extraire de l'information en renvoyant les documents électroniques qui correspondent à des besoins spécifiques;
- l'aide à la décision: les ressources linguistiques peuvent aider l'entreprise dans l'analyse automatique des textes produits par les concurrents, à travers un diagnostic rapide, et de se positionner ainsi par rapport au marché et par rapport aux concurrents;
- l'aide à la communication: la structuration linguistique des textes produits par l'entreprise peut aider cette dernière à développer un système de communication intra-organisme et extra-organisme.

Bibliographie

- Buvet (P.-A.), à paraître: *Remarques sur la détermination en français*, thèse d'habilitation, Université de Villetaneuse.
- Buvet (P.-A.) *et al.* 2005: « Les prédicats d'<affect> », dans *LIDIL*, vol. 32, Presses universitaires de Grenoble, Grenoble, p. 125-143.
- Fairon (C.), 2000: *Structures non connexes. Grammaire des incises en français: description linguistique et outils informatiques*, thèse de doctorat, Université Paris 7.
- Fellbaum (Ch.), 1999: « La représentation des verbes dans le réseau sémantique WORDNET », dans *Langages*, vol. 136, Larousse, Paris, p. 27-40.
- Hooper (P.) et Traugott (C.), 1993: *Grammaticalization*, Cambridge University Press.
- Giordano (Y.), 1995: « Communication d'entreprise: faut-il repenser les pratiques managériales? », dans *Revue de Gestion des Ressources*, vol. 13, ESKA, p. 49-61.

- Grant (D.), Oswick (C.) et Keenoy (T.), 1997 : « Organizational discourses : Text an Context », dans *Organization*, vol. 4, n° 2, p. 147-157.
- Gross (G.), à paraître : *Sémantique de la cause*, Publications de la Société de Linguistique de Paris, collection Linguistique, Peeters Louvain-Paris.
- Gross (G.), et Vivès (R.), 1986 : « Les constructions nominales et l'élaboration d'un lexique-grammaire », dans *Langue française*, vol. 69, p. 5-27.
- Gross (G.) et Prandi (M.) 2004 : *La finalité. Fondements conceptuels et genèse linguistique*, coll. Champs linguistiques, De Boeck-Duculot, Bruxelles.
- König (E.) et Traugott (E.), 1988 : « Pragmatic strengthening and semantic change: the conventionalizing of conventional implicature », dans Hüllen (W.) et Schulze (R.), éd., *Understanding the Lexicon. Meaning, Sense and World Knowledge in Lexical Semantics*, Niemeyer, Tübingen, p. 110-124.
- Massoussi (T.), 2008a : *Mécanisme de la métonymie, approche syntactico-sémantique*, thèse de doctorat, Université de Paris 13, Villetaneuse.
- Massoussi (T.), 2008b : « Représentations des inférences dans un dictionnaire électronique », dans *Cahiers de Lexicologie*, vol. 93, n° 2, p. 65-83.
- Massoussi (T.), 2009 : « Transferts sémantiques et langues spécialisées », dans *Cahiers de Lexicologie*, vol. 94, n° 1.
- Massoussi (T.) et Mejri (S.), à paraître : « Procédures d'étiquetage sémantique des textes stratégiques », *Actes du 3e colloque de doctorants et de jeunes chercheurs*, mars 2009, Université de Sousse.
- Mejri (S.), à paraître : *L'influence du discours des entreprises sur le discours des analystes*, thèse de doctorat, Université de Basse-Normandie.
- Mejri (S.), 1997 : *Le figement lexical: descriptions linguistiques et structuration sémantique*, Tunis, Publications de la faculté des Lettres de la Manouba.
- Silberztein (M.), 1993 : *Dictionnaires électroniques et analyse automatique de textes. Le système INTEX*, Masson, Paris.

Mathieu Guidère,
GRETI/LIDILEM,
Université de Genève (ETI),
Genève,
Suisse,
Mathieu.Guidere@unige.ch.

Vers un dispositif interactif d'apprentissage de la traduction à distance

Résumé

Si les outils d'apprentissage des langues à distance ont connu un essor fulgurant au cours de la dernière décennie, ceux de la traduction pâtissent d'un retard préjudiciable. Le projet DIATRAD vise le développement d'un dispositif interactif d'apprentissage de la traduction à distance en permettant aux étudiants et aux traducteurs de se former en ligne par le biais des technologies de l'internet.

Le projet est multilingue mais le dispositif est pour l'heure développé pour la traduction français-arabe et anglais-arabe. Il peut être utilisé pour une formation de type universitaire ou bien, de façon modulaire, dans la formation continue. Il s'appuie sur des séquences interactives visant à faciliter l'acquisition des principales compétences traductionnelles.

Cet article présente les principes théoriques et méthodologiques qui ont présidé à l'élaboration du dispositif, ainsi que des exemples d'activités pour la traduction français-arabe.

Mots-clés : traduction ; apprentissage ; interactif ; collaboratif ; internet ; arabe.

Introduction

L'enseignement et l'apprentissage des langues bénéficient depuis plusieurs années des innovations didactiques et technologiques liées à la généralisation de l'internet. Mais la formation à distance dans le domaine de la traduction est restée relativement à l'écart de cette évolution, en particulier pour les couples de langues éloignées telles que le français et l'arabe ou l'anglais et l'arabe.

Partant de ce constat, le projet DIATRAD (Dispositif interactif d'apprentissage de la traduction à distance), mené au sein du GRETI de l'Université de Genève (www.unige.ch/eti/GRETI), a pour objectif de s'appuyer sur les technologies et les ressources disponibles sur l'internet pour offrir aux étudiants et aux traducteurs

un environnement de travail permettant l'acquisition des principales compétences traductionnelles.

L'internet représente aujourd'hui un moteur d'innovation didactique parce qu'il associe les ressources des médias classiques avec les avantages de l'interconnexion et de l'interactivité. Le Web 2.0, en particulier, constitue un changement de paradigme, passant d'un réseau centré sur lui-même à un réseau centré sur les utilisateurs, c'est-à-dire sur l'interaction et le contenu. Ce changement de paradigme se caractérise également par le multilinguisme qui accompagne naturellement cette interaction ouverte et généralisée entre internautes issus de contrées, de langues et de cultures différentes. Dans ce contexte multilingue, la traduction apparaît comme une fonction centrale du Web et son apprentissage ne peut plus être envisagé sans recours aux outils et aux ressources offertes par l'internet à tous les utilisateurs.

1 Un dispositif centré sur les possibilités du web

Internet offre aujourd'hui un volume considérable de logiciels libres et de données exploitables pour l'enseignement et l'apprentissage de la traduction quel que soit le couple de langues considéré. Le web sémantique permet notamment de retrouver n'importe quelle information à partir de quelques fragments significatifs de celle-ci. Il permet également la constitution d'ontologies et de corpus multilingues et interconnectés. Il permet enfin le partage de ces données et la participation active des utilisateurs de différentes langues à travers des applications interactives et collaboratives.

Il existe ainsi des outils permettant de mettre en place des mécanismes de partage d'expériences et de connaissances : des groupes de discussion, des forums, des *chats*, des blogs, des galeries d'images, des agendas, des gestionnaires de flux d'informations, etc. Ces outils visent principalement à faciliter le travail en groupe à travers un espace partagé, un agenda commun, un carnet de liaison, un wiki ou encore un campus virtuel.

1.1 Une plateforme d'apprentissage en ligne

Le projet DIATRAD se caractérise par l'utilisation d'un certain nombre d'outils issus des nouvelles technologies dans le domaine de l'éducation et de l'apprentissage à distance. Il propose un environnement d'apprentissage virtuel basé sur la plateforme *Moodle* (www.moodle.org). Celle-ci offre la possibilité de mettre en place des cours en ligne dans une perspective pédagogique socioconstructiviste, c'est-à-dire privilégiant les connaissances construites à travers les interactions. Ainsi en réfléchissant sur leurs expériences de traduction, les étudiants sont amenés à construire leurs propres compétences.

La polyvalence de la plateforme *Moodle* permet de l'utiliser dans diverses activités et besoins d'enseignement. Sa structure modulaire intègre, outre des modules de base comme les cours, le calendrier et les devoirs, des modules plus spécifiques

comme les wikis, les glossaires, les forums, le *chat* et le blog (voir en annexe Fig. 1). Son utilisation peut varier d'une simple plateforme de gestion de contenu, où l'enseignant dépose les ressources des cours, à un support complet de *e-learning* dans lequel les cours sont donnés *via* la plateforme. *Moodle* se prête enfin à un enseignement hybride, associant des cours en présentiel et des activités à distance.

Du point de vue méthodologique, les fonctionnalités wiki, glossaire et forum de la plateforme permettent d'envisager des activités axées sur la recherche documentaire en intégrant la dimension collaborative. *Moodle* constitue, en effet, un puissant outil d'appui à la création et à la gestion collaborative de contenu multilingue. Ses fonctionnalités aident les traducteurs dans la résolution de difficultés de recherche en réunissant dans un même espace les ressources dont ils ont besoin, en particulier les dictionnaires et les bases de données terminologiques. Cette synergie des ressources de traduction, associée au partage de connaissances entre utilisateurs, permet un gain de temps et d'énergie non négligeable.

2 Les compétences traductionnelles au cœur du système

Il existe plusieurs méthodologies d'enseignement et d'apprentissage de la traduction, mais toutes ne conviennent pas à un apprentissage à distance. En particulier, les méthodes traditionnelles de transmission des connaissances, qu'elles soient magistrales ou mimétiques, se prêtent mal à une mise en œuvre virtuelle parce qu'elles sont fondées principalement sur le contact humain et direct entre enseignant et apprenant. À l'inverse, dans les dispositifs d'apprentissage médiatisés, l'essentiel du processus d'apprentissage est pris en charge par l'apprenant et construit en (semi-) autonomie par lui. Le rôle de l'enseignant se situe en amont, dans la phase d'élaboration des activités, phase au cours de laquelle il doit se mettre dans la peau de l'apprenant pour lui offrir un objectif et une progression susceptibles de l'aider dans son apprentissage.

Il en résulte que le dispositif, pour être efficace, doit viser des *compétences à atteindre* et non pas des *contenus à assimiler*. En traduction, ce principe est d'autant plus important qu'il existe une multitude de types textuels et une variété infinie de structures linguistiques et de vocabulaires spécialisés, qu'il est impossible de faire assimiler par un quelconque système.

De plus, envisager un *apprentissage par compétences* de la traduction à distance autorise la mise en œuvre d'une pédagogie différenciée dans laquelle chaque apprenant adapte le dispositif à ses besoins et à son rythme : il peut personnaliser son apprentissage, ce qui est une source importante de motivation.

Mais cela suppose, au préalable, la mise en place de « tests de niveaux » permettant à chaque apprenti traducteur de se situer parmi les compétences (lesquelles sont maîtrisées et lesquelles sont lacunaires ?) et sur l'échelle des compétences (quel degré de maîtrise pour chaque type de compétence ?).

2.1 Un apprentissage axé sur les compétences en traduction

Partant de ces principes, DIATRAD est basé sur un ensemble de compétences traductionnelles et contient une batterie de tests visant à évaluer le niveau de l'apprenant pour lui permettre de construire son parcours personnalisé à partir d'un diagnostic de ses propres lacunes linguistiques, culturelles, traductionnelles, méthodologiques, etc. En ce qui concerne les compétences visées, le dispositif s'appuie sur le référentiel défini dans le cadre du Master européen de traduction (EMT, *European Master in Translation*: http://ec.europa.eu/dgs/translation/programmes/emt/index_fr.htm).

Ce référentiel européen définit six grands types de compétences dont nous donnons ci-après un aperçu synthétique (tableau 1) :

| Compétence linguistique | |
|-----------------------------------|--|
| Résumé | Exercices visant à vérifier puis à consolider la maîtrise des langues en présence. |
| | 1. Comprendre les structures grammaticales, lexicales et idiomatiques ainsi que les conventions graphiques et typographiques de sa langue A et de ses autres langues de travail (B, C) |
| | 2. Utiliser ces mêmes structures et conventions en langues A et B |
| | 3. Développer sa sensibilité au changement langagier, à l'évolution des langues |
| Compétence textuelle | |
| Résumé | Exercices visant à vérifier puis à consolider la compréhension des textes à traduire. |
| | 4. Comprendre et analyser la macrostructure d'un document, sa cohérence d'ensemble (y compris quand il est composé d'éléments visuels et sonores) |
| | 5. Appréhender les présupposés, l'implicite, les allusions, les stéréotypes, l'intertextualité d'un document |
| | 6. Décrire et évaluer ses problèmes de compréhension et définir des stratégies de résolution de ces problèmes |
| | 7. Extraire et résumer les informations essentielles d'un document (capacité de synthèse) |
| | 8. Reconnaître, identifier les éléments, valeurs et référents propres aux cultures mises en présence |
| | 9. Rapprocher, comparer des éléments culturels, des modes de composition |
| | 10. Composer un document selon les conventions de genre et les normes rhétoriques |
| | 11. Rédiger, reformuler, restructurer, condenser, post-éditer vite et bien (en langues A et B) |
| Compétence interculturelle | |
| Résumé | Exercices visant à vérifier puis à consolider la maîtrise de la dimension sociolinguistique et culturelle de la traduction. |
| | 12. Reconnaître les fonctions et le sens des variations langagières (sociales, géographiques, historiques, stylistiques) |
| | 13. Identifier les règles d'interaction propres à une communauté spécifique, y compris les éléments non verbaux (savoir utile pour être apte à négocier) |

| | |
|---|--|
| 14. Produire un registre approprié à une situation donnée, pour un document (écrit) ou un discours (oral) particulier | |
| Compétence informationnelle | |
| Résumé | Exercices visant à vérifier puis à consolider la maîtrise des techniques de la documentation et de la recherche d'informations pertinentes pour la traduction. |
| 15. Identifier ses besoins en information et en documentation | |
| 16. Développer des stratégies de recherche documentaire et terminologique (y compris auprès d'experts) | |
| 17. Extraire, traiter des informations pertinentes pour une tâche donnée (informations documentaires, terminologiques, phraséologiques) | |
| 18. Développer des critères d'évaluation vis-à-vis des documents accessibles sur le Web ou tout autre support, c'est-à-dire savoir évaluer la fiabilité des sources documentaires (esprit critique) | |
| 19. Utiliser efficacement des moteurs de recherche et des outils d'aide à la traduction (exemples : logiciels de terminographie, corpus électroniques, dictionnaires électroniques) | |
| 20. Maîtriser l'archivage de ses propres documents | |
| Compétence thématique | |
| Résumé | Exercices visant à vérifier puis à consolider la maîtrise des méthodes et des techniques d'analyse du contenu thématique. |
| 21. Rechercher l'information appropriée pour mieux appréhender les aspects thématiques d'un document (cf. compétences en extraction et en exploitation de l'information) | |
| 22. Développer ses connaissances dans des domaines de spécialité et des matières d'application (maîtrise des systèmes de concepts, des modes de raisonnement, des modes de présentation, du langage contrôlé, de la terminologie, etc.) | |
| 23. Développer sa curiosité, son esprit d'analyse et de synthèse | |
| Compétence technologique | |
| Résumé | Exercices visant à vérifier puis à consolider la maîtrise des outils d'aide à la traduction, notamment des mémoires de traduction. |
| 24. Intégrer divers logiciels d'aide à la correction, à la traduction, à la terminographie, à la mise en page, à la recherche documentaire | |
| 25. Créer, gérer une base de données et de fichiers | |
| 26. Se familiariser avec de nouveaux outils, notamment pour la traduction des documents multimédias ou audiovisuels | |
| 27. Préparer et produire une traduction selon divers formats et pour divers supports techniques | |
| 28. Connaître les possibilités et les limites de la traduction automatique (TA) | |

Tableau 1

Ces compétences ont pour objectif principal l'acquisition d'une méthode de traduction en maîtrisant les différents aspects du métier. En s'entraînant à la réalisation de tâches variées correspondant à des problèmes rencontrés par les traducteurs dans les situations professionnelles, l'apprenant construit son parcours suivant des

objectifs précis et à partir de ses propres besoins et lacunes (que sais-je [faire]? que dois-je améliorer?).

3 Les textes comme support privilégié

Pour acquérir une méthode de traduction finalisée (c.-à-d. fondée sur des objectifs), le dispositif offre une approche textuelle des problèmes de traduction en fonction des couples de langues. En d'autres termes, les exercices ne sont pas conçus à partir de collocations ni de phrases isolées ou pris hors de leur contexte, mais à partir de textes authentiques qui présentent ces difficultés dans leur environnement verbal et culturel.

De ce point de vue, DIATRAD propose aux traducteurs deux types d'activités textuelles : d'une part, des *activités de production* dans lesquelles l'apprenant traduit, à partir des textes qui lui sont présentés, certains segments ou séquences textuelles problématiques ; d'autre part, des *activités de révision* dans lesquelles l'apprenant valide, à partir de productions déjà existantes, des choix de traduction qui nécessitent une justification.

Le dispositif ne présente pas à l'apprenant des choix de traduction erronés pour éviter tout risque de confusion ou de fossilisation des erreurs. Toutes les possibilités présentées sont correctes du point de vue linguistique, mais leur degré d'adéquation au contexte (sur le plan culturel, pragmatique, stylistique, etc.) mérite en général une réflexion et une validation de la part de l'utilisateur. En tout état de cause, il s'agit de traductions réalisées par des professionnels ou des enseignants de traduction ayant une visée didactique avérée.

3.1 Un apprentissage tourné vers les textes de spécialité

Pour une meilleure structuration de l'apprentissage, le travail sur les compétences est tourné vers des textes issus de domaines spécialisés bien identifiés. La raison en est que certaines compétences traductionnelles sont plus facilement exploitables dans certains types de textes. Par exemple, il est plus intéressant de travailler la compétence interculturelle à partir de textes littéraires qu'à partir de textes techniques. À l'inverse, les textes techniques et scientifiques sont plus appropriés pour le développement de la compétence documentaire et terminologique.

C'est pourquoi nous avons défini, dans le cadre de DIATRAD, cinq types de textes professionnels, qui préparent aux domaines suivants de formation : la traduction économique (à partir de textes sur l'économie et la finance), la traduction juridique (à partir de textes de droit national et international), la traduction médicale (à partir de textes concernant la santé, la médecine et la pharmacie), la traduction technique (à partir de textes principalement informatiques), la traduction publicitaire (à partir de textes promotionnels et de sites web commerciaux). Chacun de ces types de textes pose des problèmes spécifiques mais mobilise des compétences communes.

3.2 Un apprentissage par problème

À partir de cette variété de textes, nous avons défini, pour chaque couple de langues, un ensemble de problèmes de traduction que l'utilisateur doit apprendre à identifier et à traiter en se forgeant une méthode de travail à partir des ressources disponibles.

Nous donnons ci-après un aperçu des problèmes traductionnels et des ressources disponibles pour le couple français-arabe (tableau 2) :

| Problèmes de traduction (FR/AR) | |
|---|---|
| Résumé | Exercices visant à vérifier puis à consolider la maîtrise des langues en présence. |
| 1. Les problèmes linguistiques : les temps verbaux et la concordance des temps ; les types de phrases ; les adverbes, les circonstants et les mots outils. | |
| 2. Les problèmes terminologiques : les termes ; les synonymes ; les termes polysémiques ; les domaines et les champs d'activité et de pratique. | |
| 3. Les problèmes culturels : les expressions idiomatiques et les collocations ; les contenus politiques ; les contenus religieux ; les références littéraires et les métaphores poétiques. | |
| Activités associées (FR/AR) | |
| Résumé | Exercices visant à résoudre les problèmes diagnostiqués par l'apprenant à chaque étape. |
| 4. L'identification et le traitement des problèmes linguistiques : la traduction des temps verbaux ; la traduction des adverbes et des mots outils ; la traduction des différents types de phrases. | |
| 5. L'identification et le traitement des problèmes terminologiques : la traduction des synonymes ; la traduction des termes polysémiques ; la traduction des expressions idiomatiques et des collocations. | |
| 6. L'identification et le traitement des problèmes culturels : la traduction des contenus politiques et idéologiques ; la traduction des références religieuses ; la traduction des références littéraires et des métaphores poétiques. | |
| Ressources associées (FR/AR) | |
| Résumé | Ressources lexicales et informationnelles visant la maîtrise de la dimension documentaire et interdisciplinaire de la traduction. |
| 7. Le vocabulaire politique (FR/AR) : glossaire bilingue de politique nationale et internationale ; dictionnaires politiques en ligne. | |
| 8. Le vocabulaire économique (FR/AR) : glossaire bilingue de l'économie et de la finance ; dictionnaires économiques en ligne. | |
| 9. Le vocabulaire juridique (FR/AR) : glossaire bilingue du droit musulman ; dictionnaires juridiques en ligne. | |
| 10. Le vocabulaire médical (FR/AR) : glossaire bilingue de la santé, de la médecine et de la pharmacie ; dictionnaires médicaux en ligne. | |
| 11. Le vocabulaire technique (FR/AR) : Glossaire bilingue de l'informatique, de la physique et de la mécanique ; dictionnaires techniques en ligne. | |

| Bibliographie indicative en traduction (FR/AR) : introduction à la traductologie ; manuels de traduction ; articles en ligne sur la traduction français-arabe ; revues en ligne ; bases de données en ligne ; sites web de traduction automatique ; sites web des médias et de la presse française et francophone ; sites web des médias et de la presse arabe et arabophone. | |
|---|---|
| Types de tests (FR/AR) | |
| Résumé | Tests visant à vérifier le niveau de l'apprenant à l'entrée et à la sortie, pour chaque type de compétence. |
| 13. Questions à choix multiples (QCM) sur des possibilités de traduction. | |
| 14. Questions « à trous » sur des choix terminologiques. | |
| 15. Questions à compléter avec des propositions personnelles de traduction. | |
| 16. Questions « à relier » avec des équivalences idiomatiques. | |
| 17. Questions « à cocher » avec des solutions ou choix alternatifs. | |
| 18. Questions « à rédiger » à partir de segments ou de séquences textuelles. | |

Tableau 2

L'objectif de ce parcours est d'évaluer le niveau de l'étudiant avant de se lancer dans la formation en autonomie. Les différents tests ont été spécialement conçus pour des étudiants francophones et arabophones : ils visent la familiarisation avec les techniques de base de la traduction professionnelle et avec les problématiques les plus récurrentes dans la traduction du français vers l'arabe et inversement.

Il est à noter que la réaction des étudiants face à ces activités est très positive : ils apprécient le caractère à la fois « autonome » et « guidé » du dispositif. On remarque même un esprit d'émulation entre eux qui permet d'optimiser les bénéfices du parcours. En tout état de cause, les premiers résultats montrent un intérêt indéniable pour le dispositif et une contribution à la conscientisation des étudiants face à leur processus d'apprentissage.

Conclusion

Bien que proposant de nombreuses séquences et ressources, DIATRAD ne constitue pas une formation à la traduction générale. Il est conçu comme un complément à une formation à la traduction spécialisée dans les domaines les plus recherchés sur le marché du travail. En effet, les modules disponibles à ce jour abordent des types de textes variés allant de la médecine à l'informatique en passant par l'économie et le droit. Ces séquences thématiques permettent à l'étudiant de travailler les principales compétences traductionnelles de façon interactive, en fonction de ses centres d'intérêt et de ses lacunes personnelles dans tel ou tel champ du savoir.

Les activités sont conçues pour un travail en autonomie en complément de la formation classique à la traduction, mais elles peuvent également être intégrées par les enseignants dans les cours de traduction existants. Le dispositif visait initialement la formation continue des traducteurs arabophones qui ont le français ou

l'anglais dans leur combinaison linguistique, mais les résultats des usages observés montrent qu'il peut contribuer de façon significative à l'acquisition des langues de spécialité par tous les étudiants en traduction.

Il est important de noter, en effet, que les principes, les méthodes et les techniques employés dans DIATRAD sont applicables à la traduction dans d'autres couples de langues. Ils exigent néanmoins un important travail de conceptualisation didactique en amont des applications. Ce travail ne peut être mené que par des enseignants de traduction sensibilisés aux nouvelles technologies éducatives et à l'apport spécifique de l'internet dans la formation en langues et en traduction. En aucun cas, il ne peut s'agir de mettre en ligne des cours de traduction « classiques » conçus initialement pour une formation en présentiel.

Mais pour rassurer les esprits réticents à l'égard de la technologie, il est utile de rappeler que l'accompagnement pédagogique des étudiants sur les modules décrits plus haut ne nécessite pas un haut degré de qualification technique pour l'enseignant en charge du cours. Plus que l'obstacle technologique, le formateur concepteur de tels modules doit surmonter la barrière psychologique qui l'empêche parfois d'innover pour avancer.

Annexes

Dans ces annexes, on trouvera des copies d'écrans des types d'activités proposées dans le cadre de DIATRAD avec une description succincte des protocoles.

| <i>Analyse de la langue A en vue de la traduction</i> |
|---|
| <p>Objectifs: Permettre aux apprenants de développer leur sensibilité aux variations langagières (sociales, géographiques, historiques, stylistiques) en vue de la traduction.</p> <p>Descriptif: Le module se propose de compléter les connaissances acquises en apprentissage de la langue A, en se focalisant sur la question des variations langagières.</p> <p>Compétences visées: Savoir analyser la langue A du point de vue pragmatique: savoir identifier et comprendre des variétés langagières, des genres textuels, des registres et styles différents de la langue A; connaître les différents systèmes d'équivalence qu'offre la Langue A. (Selon les besoins de chaque apprenant)</p> |
| Évaluation formative; exercices pratiques. |
| Travail en autonomie. |

Figure 1 : fiche signalétique d'une activité sous DIATRAD

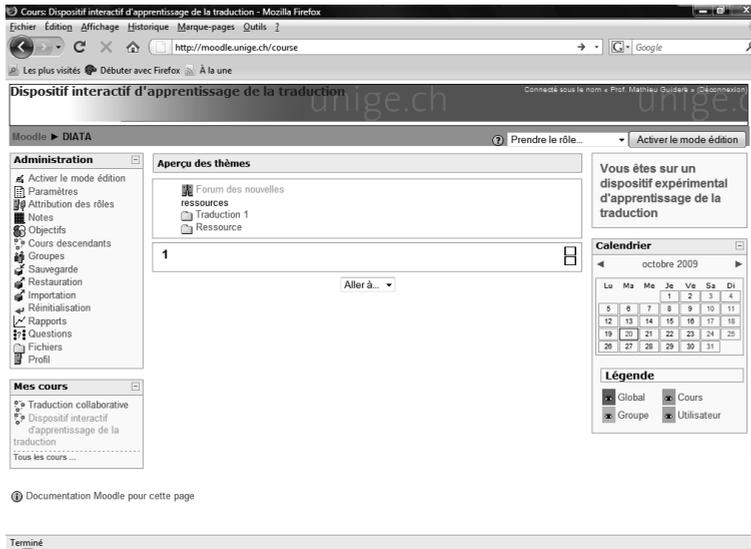


Fig. 2 : Page d'accueil de DIATRAD sur la plateforme Moodle

Ci-après la description des différentes activités qui sont offertes à l'apprenant dans le module de traduction *de* et *vers* l'arabe.

Protocole 1 : traduire des termes en contexte (à partir d'un texte)



Fig. 3 : exemple issu d'un texte informatique

Le protocole d'exécution est le suivant : l'apprenant lit le texte puis relie le terme indiqué à l'équivalent approprié. Il déplace les éléments et les met vis-à-vis. Une

fois ses choix effectués, il clique sur le bouton de validation pour vérifier la pertinence de ses réponses. Un score est affiché.



Fig. 4 : Exemple issu d'un texte juridique

Le protocole d'exécution est le suivant : l'apprenant lit le texte puis relie le terme indiqué à l'équivalent approprié. Il déplace les éléments et les met vis-à-vis. Une fois ses choix effectués, il clique sur le bouton de validation pour vérifier la pertinence de ses réponses. Un score est affiché.

Protocole 2 : Traduire des expressions idiomatiques

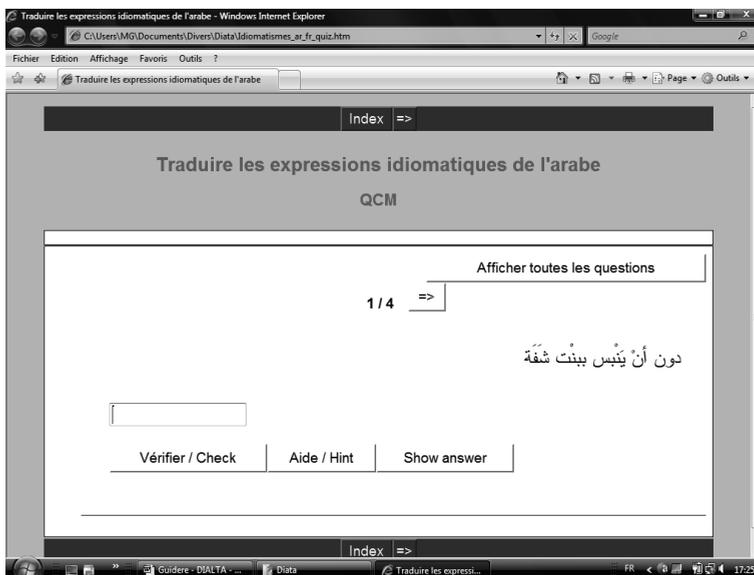


Fig. 5: Exemple à partir de l'arabe

Quiz textuel : L'apprenant essaie de trouver l'équivalent adéquat à partir d'un texte.

Le protocole d'exécution est le suivant : l'apprenant lit la phrase puis propose une traduction. Il saisit sa proposition dans l'espace prévu à cet effet. Une fois sa proposition de traduction saisie, il clique sur le bouton de validation (*Vérifier/Check*) pour vérifier la pertinence de sa traduction.

Il a également la possibilité d'afficher la bonne solution (*Show answer*) ainsi que le commentaire associé avant de passer à l'écran suivant. Le passage d'un exercice à l'autre se fait suivant un mode d'auto-entraînement, avec indication d'un score de passage.



Fig. 6: exemple français-arabe (expressions idiomatiques)

QCM textuel: L'apprenant relie l'expression à la traduction appropriée.

Le protocole d'exécution est le suivant: l'apprenant lit le texte puis relie le terme indiqué à l'équivalent approprié. Il déplace les éléments et les met vis-à-vis. Une fois ses choix effectués, il clique sur le bouton de validation pour vérifier la pertinence de ses réponses.

Protocole 3: traduire des problèmes linguistiques contextualisés

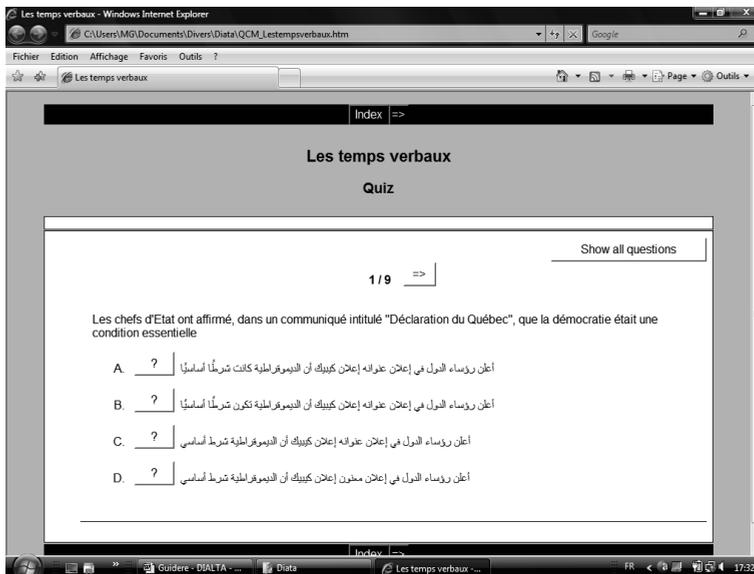


Fig. 7: traduire les temps verbaux

QCM textuel à partir d'un discours : l'apprenant doit trouver la bonne traduction de la phrase à partir d'un texte.

Le protocole d'exécution est le suivant : l'apprenant lit l'énoncé à traduire puis coche la case correspondant à la traduction qu'il juge appropriée. Dès que son choix est fait, l'ordinateur affiche une réponse (vrai/faux), avec un commentaire concernant la validité de la proposition. L'apprenant a également la possibilité d'afficher la bonne solution (*Show answer*), ainsi que le commentaire associé à toutes les réponses avant de passer à l'écran suivant. Le passage d'un exercice à l'autre se fait suivant un mode d'auto-entraînement, avec indication d'un score de passage.

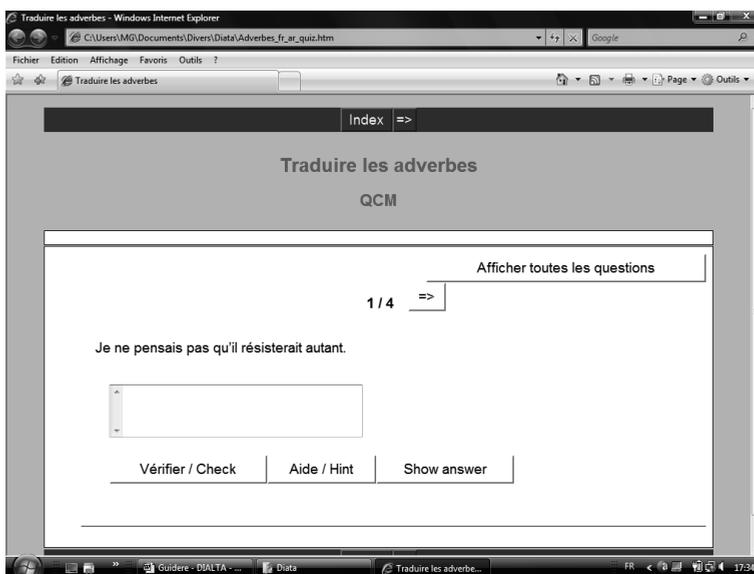


Fig 8 : quiz sur les adverbes (p ex. : « autant »)

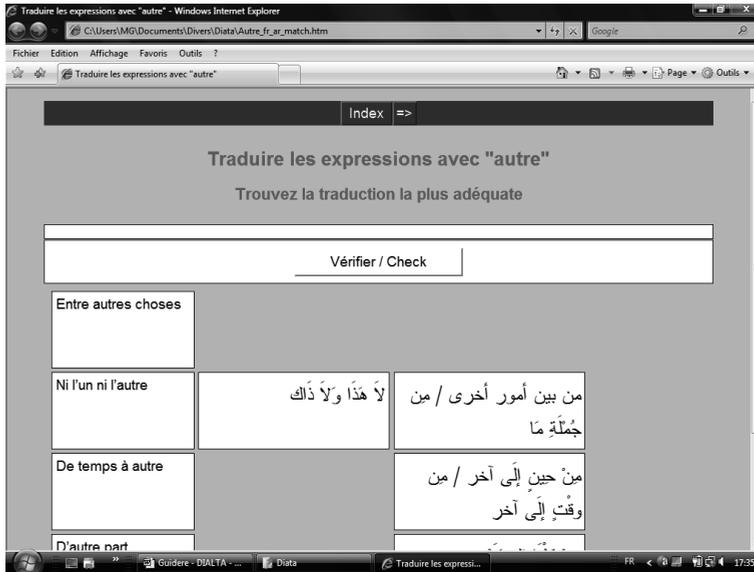


Fig. 9: traduire les expressions françaises avec « autre »

Concordance textuelle : l'apprenant doit associer la bonne expression à sa traduction.

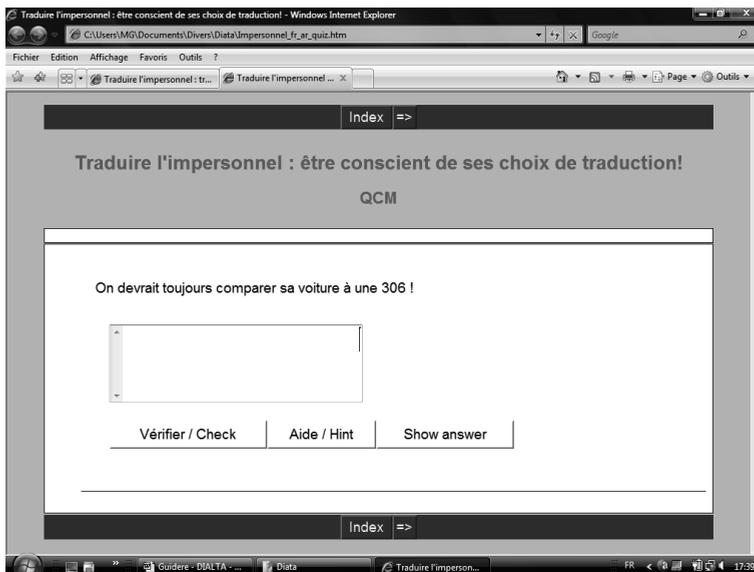


Fig. 10: Traduire l'impersonnel du français vers l'arabe

L'apprenant doit trouver la traduction appropriée à une phrase contenant l'impersonnel (*on*).

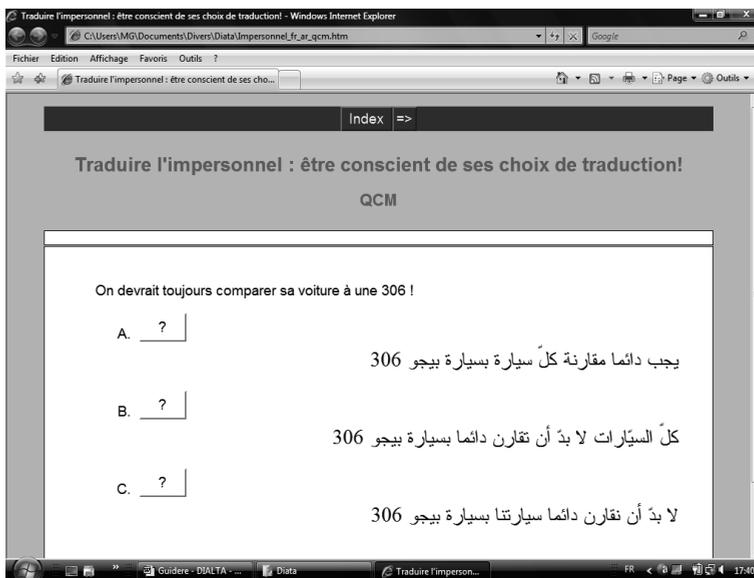


Fig. 11: traduire les slogans publicitaires (reformulation et justification)

QCM textuel: l'apprenant doit indiquer la traduction la plus précise pour un énoncé.

Protocole 4: Traduire des textes spécialisés

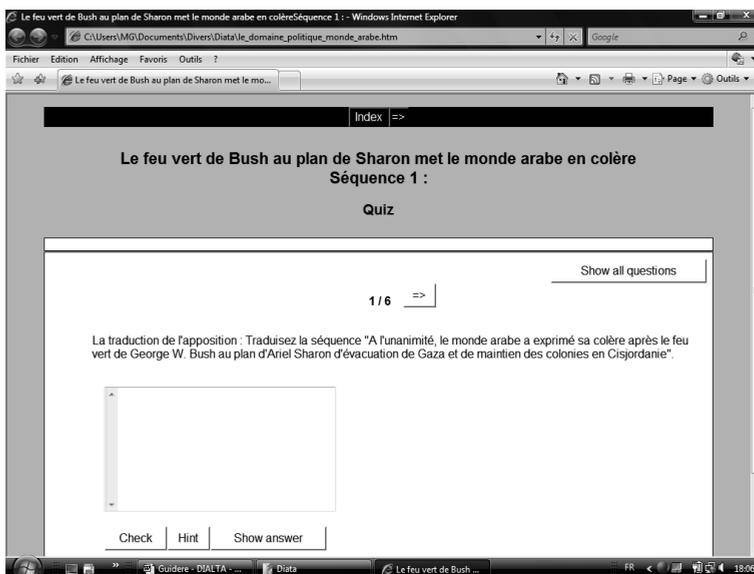


Fig. 12: Traduire un texte relevant du domaine politique

L'apprenant doit traduire l'énoncé en tenant compte du problème indiqué (ici la traduction de l'apposition).

Hong Kong en comparaison affiche une performance décevante (moins 10 % depuis le début de l'année), déséquilibrée par quatre sociétés, HSBC, China Mobile, Hutchison Whampoa et Cheung Kong Holdings, qui représentent 62 % de l'indice Hang Seng. « China Mobile, que les investisseurs ont longtemps conservé en dépit de la baisse du secteur des télécoms, subit aujourd'hui des ventes de capitulation », explique le spécialiste. Il n'est pas certain que de tels mouvements de vente ne touchent à leur tour Taipei ou Séoul, si la croissance américaine s'avère incapable de relancer la demande de composants électroniques, pilier de ces économies.

La Tribune du 16/03/2001

Quiz

Show all questions

1 / 3 =>

La traduction des chiffres : choisissez la bonne traduction du paragraphe allant de "Taipei a fini en hausse ... du marché chinois coté en devises".

A. أُنقِطت سوق تايبي على 5.742.74 أي بارتفاع نسبتته 1.5 % أما سيول وانقُطت على 541.83 نقطة أي بانخفاض نسبتته 0.3 % - على انقطت سوق تايبي على 5.742.74 أي بارتفاع نسبتته 1.5 % أما سيول وانقُطت على 541.83 نقطة أي بانخفاض نسبتته 0.3 % بالمالحة

Fig. 13 : Traduire un texte relevant du domaine économique

QCM textuel : l'apprenant doit choisir la traduction adéquate au paragraphe indiqué.

Le domaine médical : Nouveau pas vers un vaccin tueur de tumeurs

Quiz

Show all questions

1 / 2 =>

Traduction de la probabilité : Traduisez la séquence suivante "Rêvons à un âge ultérieur. En ce temps-là, il y aurait, contre les cancers, un traitement simple et efficace. Une simple piqûre. Le médecin injecterait au malade une molécule chimique qui agirait comme le bon vieux vaccin antirabique. La substance aurait les vertus de déclencher contre la source du mal – en l'occurrence, les cellules cancéreuses – les foudres du système immunitaire. Celui-ci, sous l'effet du produit injecté, saurait soudain reconnaître, entre toutes les cellules, les malignes à abattre. On appellerait ce traitement un vaccin anti-cancer. Vers ce rêve auquel travaillent des scientifiques depuis une bonne dizaine d'années, un nouveau pas vient d'être franchi."

A. لنظم بزمن قادم ولكن فيه علاج سهل ومعالج ضد السرطان ولكن سره في حقنة بسيطة بحقن الطبيب بواسطة حقنة كيميائية مفعولها مثل مفعول القاح ضد الكلب الشهير. فتسمح هذه المادة لنظام المناعة بترجوه بصرته ضد مصدر المرض أي الخلايا السرطانية إذ يستغلون بفعال المادة المحفزة التعرف على الخلايا البنيوية التي يجب القضاء عليها من بين كل الخلايا. ويسمى هذا العلاج القاح ضد السرطان لأنه يحفز الجهاز المناعي على مهاجمة الخلايا السرطانية وحدها.

B. لنظم بمر قادم فيكون فيه ضد السرطان علاج سهل ومعالج بواسطة حقنة بسيطة. وقد يحقن الطبيب المريض حقنة كيميائية فيكون مفعولها مثل القاح ضد الكلب القديم وقد يكون لهذه المادة البزوايا لأن تحفز سواقي نظام المناعة ضد مصدر الضرر في هذه الحالة الخلايا السرطانية. وهذا قد يستغل تحت تأثير المادة المحفزة التعرف من بين كل الخلايا على الخلايا التي يجب قتلها وقد يسمى هذا العلاج قاحاً ضد السرطان. نعر هذا الخطر إلى الأمام بجداء تحقيق هذا الحلم الذي يطمنون عليه منذ أكثر من عشر سنوات.

Fig. 14 : traduire un texte relevant du domaine médical

QCM textuel : l'apprenant doit vérifier la traduction du paragraphe indiqué à partir d'un ensemble de propositions dont une seule est en adéquation avec le contexte.

Abdallah El Mountassir,
Université Ibn Zohr, Département de français, Filière d'études amazighes,
Agadir,
Maroc,
elmountassir@hotmail.com.

Vers un modèle d'intercompréhension progressive entre l'amazigh et le français. L'expérience de l'enseignement de l'amazigh à l'université marocaine

Résumé

L'expérience récente de l'enseignement de l'amazigh à l'université marocaine se heurte à quelques difficultés liées à l'absence d'une terminologie grammaticale et d'un métalangage propres à cette langue. Dans le contexte actuel, les cours d'amazigh sont donnés soit en langue française (70 % des matières enseignées), soit en langue arabe, cela dépend de la formation des enseignants et des matières enseignées. Dans cet article, nous abordons l'expérience de cet enseignement à l'université d'Agadir (Maroc). Nous nous intéressons au rôle de la langue française dans l'enseignement de l'amazigh en nous basant sur les aspects suivants: langue d'enseignement, question de terminologie grammaticale et développement d'une intercompréhension entre l'amazigh et le français.

Mots-clés: enseignement des langues peu documentées; transmission des connaissances locales; intercompréhension entre amazigh et français.

1 Enseignement d'une langue peu documentée en langues dominantes

1.1 Contexte de la problématique

De nos jours, on dénombre à peine 200 langues écrites et bénéficiant d'un enseignement. Il y a beaucoup de situations dans le monde où l'école constitue un système sélectif et élitiste dans la mesure où les locuteurs des langues vernaculaires sont scolarisés et alphabétisés uniquement en langues dominantes ou en langues nationales parlées par la majorité de la population, plutôt que dans leur propre langue (Boudreau *et al.* 2002).

Selon les experts, éduquer les enfants des communautés linguistiques minoritaires uniquement dans les langues dominantes présente de nombreux inconvénients pour ces communautés (échec scolaire et personnel, dévalorisation de la culture, perte du savoir et des traditions, risque de disparition des langues et des cultures, etc.).

Actuellement, nombreuses sont les communautés linguistiques minoritaires qui revendiquent le droit à leurs particularismes identitaires, et la pression sociale pour l'enseignement de leurs langues devient de plus en plus forte. L'enseignement ne constitue pas uniquement un outil important de préservation de ces langues, mais c'est aussi un moyen d'assurer leur survie.

Dans ce contexte, les linguistes et les spécialistes dans le domaine de l'éducation se heurtent à des difficultés sérieuses pour l'intégration de ces langues de tradition orale dans le système scolaire. Ces difficultés sont souvent liées au problème de la gestion des variations dialectales (élaboration d'une langue standard, norme à enseigner, normalisation du système d'écriture) et à l'absence d'outils pédagogiques de base (grammaires, dictionnaires, méthodes de langues, manuels scolaires, etc.) indispensables à l'enseignement de ces langues.

À cela, il faudrait ajouter également que la plupart de ces langues, qui sont peu documentées, n'ont pas développé une terminologie grammaticale, et que les travaux de description linguistique réalisés sur ces langues sont souvent rédigés en langues dominantes.

L'intégration de ces langues minoritaires dans le système scolaire soulève donc plusieurs questions : comment peut-on élaborer du matériel didactique dans ces langues ? Dans quelle langue devrait-on rédiger les manuels scolaires et les grammaires pour soutenir leur enseignement ? Quelle terminologie grammaticale devrait-on choisir ? Selon quels critères devrait-on créer une terminologie grammaticale pour ces langues ? Comment peut-on développer une langue de scolarisation ? Comment mettre en place un programme de formation des enseignants ?

Dans le contexte actuel, il est donc difficile et irréaliste pour beaucoup de langues vernaculaires d'assurer un enseignement uniquement en langue vernaculaire, faute de ressources humaines et matérielles appropriées. Les chercheurs et les responsables du système éducatif dans plusieurs communautés se sentent donc interpellés par ces difficultés et se trouvent devant des situations d'urgence auxquelles ils doivent réagir. Pour répondre aux besoins immédiats, certains linguistes et quelques gouvernements optent pour un enseignement des langues vernaculaires en langues dominantes. Autrement dit, enseigner les langues vernaculaires avec une terminologie grammaticale empruntée aux langues dominantes.

Pour ces linguistes, la mise en place de ce type d'enseignement devrait permettre aux locuteurs des langues vernaculaires l'apprentissage et également l'acquisition du fonctionnement grammatical de leur propre langue. Pour certains chercheurs et pédagogues, cette éducation est une solution plus réaliste pour la plupart des langues vernaculaires et minoritaires en situation d'urgence, dans la mesure où ce type d'enseignement est un moyen parmi d'autres permettant de consolider et de préserver ces langues menacées de disparition.

Il s'agit donc dans ce contexte éducatif de mettre les langues dominantes au service des langues peu documentées.

2 Enseignement de l'amazigh au Maroc

2.1 L'expérience de l'école primaire

À la suite d'un mouvement social de plus en plus fort, nous assistons à quelques changements survenus dans la politique linguistique du Maroc. Ces changements se sont traduits par une reconnaissance officielle de la langue et de la culture amazighes par l'État marocain. La langue amazighe a donc acquis une certaine légitimité, alors qu'il y a encore une dizaine d'années, la langue et la culture amazighes étaient complètement bannies du discours officiel de ce pays. L'année scolaire 2003-2004 a connu un événement important dans l'enseignement du Maroc : il s'agit de l'intégration de l'amazigh dans le système éducatif marocain. C'est la première fois dans l'histoire du Maroc que l'amazigh est enseigné, d'une manière officielle, dans certaines écoles (niveau primaire) du pays.

Cette ouverture sur la langue et la culture amazighes a des répercussions sociales et politiques au Maroc et un impact sur l'avenir de cette langue. Ce récent regain d'intérêt officiel pour l'amazigh et ce nouveau contexte de changement de statut que connaît actuellement cette langue (d'une langue dominée à une langue reconnue) a ouvert également de nouvelles perspectives de recherches et d'action au Maroc et à l'étranger.

2.2 L'intégration des études amazighes à l'université

Parmi ces perspectives d'action au Maroc, la création, pour la première fois, de filières d'études amazighes dans certaines universités du pays. Depuis l'année 2007-2008, trois filières d'études amazighes ont vu le jour. Ces filières sont localisées dans trois centres universitaires : la Faculté des lettres d'Agadir, celle de Fès et celle d'Oujda. Il s'agit des filières de la licence fondamentale en études amazighes (langue et littérature amazighes). Ces filières, qui sont ouvertes aux titulaires d'un baccalauréat en lettres, proposent un programme de formation de six semestres (trois ans) avec des modules fondamentaux majeurs (langues et littératures amazighes) et d'autres modules optionnels (histoire, sociologie, langues étrangères, techniques d'expression et de communication).

Le projet de la création des filières de licence en études amazighes intervient dans un contexte d'urgence qui est fortement marqué par, d'une part, l'accentuation de l'attente sociale à propos de l'enseignement de la langue amazighe, et, d'autre part, l'intégration de plus en plus poussée des études amazighes dans le champ de la recherche nationale et internationale.

Malgré cette ouverture sur la langue et la culture amazighes, il convient de rappeler que, dans l'état actuel, cette langue se caractérise par la variation dialectale et ne

dispose pas encore d'un système d'écriture standardisé qui est connu et maîtrisé par tous les locuteurs. Jusqu'à présent, nous n'avons pas une langue amazighe commune standard. Les langues de l'écrit, pour les locuteurs amazighophones, ce sont d'abord l'arabe (classique) et le français; ce sont les langues des documents officiels, de l'école et de la formation scientifique. Ces deux langues, qui possèdent déjà un métalangage et une tradition grammaticale bien développés, constituent donc les langues de scolarisation pour ces locuteurs. Par ailleurs, la langue amazighe souffre du manque d'outils de référence de base (grammaires, dictionnaires, manuels, etc.) rédigés dans cette langue. Toutes les descriptions grammaticales de l'amazigh sont rédigées en langues dominantes (en grande partie en français et en arabe écrit). Il faudrait rappeler, enfin, qu'il n'existe pas encore une terminologie et un métalangage propres à la langue amazighe.

Dans le contexte actuel, les cours d'amazigh sont donc donnés soit en langue française (70 % des matières enseignées), soit en langue arabe, cela dépend de la formation des enseignants et des matières enseignées.

Nous parlerons ici de l'expérience récente de l'enseignement de l'amazigh à la Faculté des lettres et des sciences humaines de l'Université d'Agadir (sud-ouest du Maroc). Nous examinerons quelques questions et difficultés concernant cet enseignement. Plus précisément, nous nous intéresserons ici au rôle de la langue française dans l'enseignement de l'amazigh en nous basant sur les points essentiels suivants :

- langue d'enseignement;
- métalangage et terminologie grammaticale;
- développement d'une intercompréhension entre l'amazigh et le français.

3 Pour une approche communicative entre l'amazigh et le français

L'objectif prioritaire, dans l'enseignement de l'amazigh en français, c'est de permettre aux étudiants l'acquisition des connaissances grammaticales et d'un savoir métalinguistique propres à la langue amazighe. Rappelons que ces étudiants ont déjà quelques connaissances grammaticales du français, acquises au cours de leurs études secondaires, mais pas en langue amazighe¹.

Dans ce contexte, il importe de préciser que toutes les connaissances métalinguistiques du français ne sont pas forcément transférables dans la langue amazighe. D'où la question incontournable à laquelle nous avons dû faire face dès le début de cette formation : jusqu'où peut-on réussir à enseigner l'amazigh avec une terminologie et un métalangage du français ?

¹ Les spécialistes parlent dans ce cas d'un processus de grammatisation qui permet aux apprenants d'acquérir un savoir métalinguistique de la langue. Par ailleurs, on oppose le savoir métalinguistique, c'est-à-dire cette réflexion consciente des locuteurs sur le fonctionnement grammatical d'une langue, au savoir épilinguistique qui est de nature intuitive (cf. bibliographie).

Cette question revient souvent dans le contexte où on enseigne une langue minoritaire en langue de grande diffusion. Il s'agit de savoir comment arriver à enseigner ces langues minoritaires tout en évitant de transposer dans ces langues les concepts grammaticaux propres aux langues dominantes.

Dans notre expérience de l'enseignement de l'amazigh à l'université, nous adoptons une approche comparative: il s'agit d'identifier les concepts grammaticaux de l'amazigh en les comparant aux équivalents du français. Cette façon de faire permet aux étudiants de bien appréhender les similitudes et les différences entre les deux langues. Une catégorie de mots en français ne correspond pas nécessairement à la même catégorie en amazigh. Là où nous avons des adjectifs en français par exemple, nous avons des verbes descriptifs en amazigh²:

[1] *y-uḍen* (= il — avoir mal) « il est malade »

[2] *t-mlul tgmimi* (= elle — être blanc maison) « la maison est blanche »

[3] *fulki-nt tg^omma-ad* (= être beau — elles maisons — ci) « ces maisons sont belles ».

Les mots *uḍen* « avoir mal, souffrir », *mlul* « être blanc » et *fulki* « être beau » sont des verbes qui se conjuguent comme n'importe quel autre verbe à toutes les personnes.

Ces exemples font partie des exercices de traduction que nous proposons aux étudiants à des fins pédagogiques dans la mesure où ils fonctionnent comme des mises en garde: ne pas parler de la catégorie de l'adjectif en amazigh. Ce type de comparaison permettra aux étudiants d'établir des parallèles entre les catégories de mots des deux langues. Cette approche comparative constitue donc, dans ce contexte, un outil pertinent pour la compréhension des particularités grammaticales de l'amazigh. Au niveau de la méthode de travail, nous rappelons que ces comparaisons entre les deux langues s'effectuent toujours à partir de textes variés et authentiques.

Dans ce qui suit, nous examinons d'autres exemples que nous jugeons pertinents pour l'élaboration des champs d'intercompréhension entre l'amazigh et le français.

4 Traduction et intercompréhension entre l'amazigh et le français

4.1 Morphologie verbale

Le système verbal en amazigh est un système aspectuel et non pas temporel comme en français. D'une manière générale, on distingue en amazigh deux formes morphologiques du verbe qui s'opposent: l'accompli et l'inaccompli. Et cette opposition ne recouvre pas une distinction temporelle mais une opposition aspectuelle. La forme verbale d'accompli envisage un procès comme achevé et terminé. La forme

² Comme c'est le cas pour bon nombre de langues peu documentées, le verbe occupe une place importante dans la catégorie de mots en amazigh.

d'inaccompli peut avoir, selon le contexte, la valeur d'un duratif (durée) ou d'un itératif (répétition du procès).

On ne parle donc pas en amazigh du temps passé/présent/futur. Dans cette langue, le temps peut être exprimé par des indices contextuels ou par d'autres termes; la forme verbale en elle-même n'indique jamais le temps.

Comment cela se traduit-il en français ?

- Une forme verbale d'inaccompli comme *ar i-ctta* dans l'exemple [4] *ar i-ctta iselman* se traduit en français par :
 - « il mange du poisson » (quelqu'un qui a l'habitude de manger du poisson);
 - « il est en train de manger du poisson » (durée au moment présent);
 - « il mangeait du poisson » (habitude dans le passé);
 - « il était en train de manger du poisson » (durée dans le passé).

Nous avons donc ici quatre sens différents (au niveau du temps) en français pour la même forme verbale amazighe *ar ic̣tta*. Chaque sens dépend d'un contexte particulier. Ces exemples illustrent bien la différence entre le système verbal aspectuel de l'amazigh et le système verbal temporel du français. Et la compréhension de cette différence (par la comparaison) des deux systèmes verbaux des deux langues permet aux étudiants de bien saisir les correspondances qu'il y a entre les deux langues. Pour une forme verbale aspectuelle de l'amazigh, l'étudiant devrait choisir une forme verbale temporelle (présent, imparfait) adéquate en français.

L'objectif ici est de développer un espace de communication et un répertoire langagier dans le domaine du système verbal des deux langues permettant à l'étudiant de passer efficacement d'une langue à l'autre.

4.2 Structure de la phrase

En amazigh, dans toute forme verbale il y a obligatoirement association d'un indice de personne et d'un radical verbal.

[5] *i-fta* « il est parti » (= *i*: indice de 3^e pers. masc. sing.; *fta*: radical)

[6] *fta-nt* « elles sont parties » (= *nt*: indice de 3^e pers. fém. pl.; *fta*: radical)

Ainsi, l'indice de personne et le radical verbal ne peuvent apparaître séparément si on veut avoir un énoncé complet. Si on veut préciser le référent de l'indice de personne *i-*, par exemple, on ajoute à l'énoncé *i-fta* un nom, qui vient après le verbe. Cette fois-ci, il s'agit d'un élément facultatif:

[7] *i-fta urgaz* « l'homme est parti » (litt. « il est parti l'homme »)

- Cette description comparative entre l'amazigh et le français permet d'attirer l'attention des étudiants sur les points suivants :
- l'indice de personne *i-* n'est pas un pronom « il » comme en français, puisqu'il s'agit d'un élément obligatoire dans l'énoncé ;
- d'un point de vue sémantique, le terme *urgaz* « homme » est un référent lexical de l'indice de personne *i-* ;
- le concept de sujet en amazigh est différent de celui du français ;
- la différence entre les deux langues au niveau de l'ordre des mots : indice de personne – verbe — référent lexical — objet, pour l'amazigh, et sujet — verbe — objet, pour le français.

Ce type de comparaison entre les deux langues permet également de mettre en garde l'étudiant contre les erreurs en français de type « *l'homme il est parti ».

5 Un métalangage français pour l'enseignement de l'amazigh

Cette approche comparative entre l'amazigh et le français nous incite à réfléchir actuellement sur l'élaboration d'une terminologie et d'un métalangage pour l'enseignement de l'amazigh en français. Comme nous l'avons précisé plus haut, plusieurs termes et concepts grammaticaux propres à la langue française ne sont pas transférables vers l'amazigh (adjectif, article, participe, etc.). Il faudrait donc choisir une terminologie qui correspondrait à la réalité de la langue amazighe. Cette option est inévitable, dans l'état actuel des choses, dans la mesure où elle permet à nos étudiants de ne pas se référer toujours à leurs connaissances métalinguistiques du français au cours de l'acquisition des connaissances linguistiques de l'amazigh. Dans le domaine de la morphologie verbale par exemple, l'amazigh, contrairement au français, n'a pas d'infinitif. Pour faire référence à un verbe en amazigh, nous avons choisi la forme verbale la plus simple qui correspond à la 2^e personne du singulier de l'impératif.

Par ailleurs, lorsqu'on enseigne une langue minoritaire en langue de grande diffusion se pose également la question du niveau de langue d'enseignement. Dans notre cas, la consigne qui a été donnée aux enseignants est l'emploi d'une langue simple, claire et précise, accessible aux étudiants. Pour cela, nous évitons l'usage d'un vocabulaire technique spécialisé ou l'emploi des concepts grammaticaux complexes qui ne sont pas familiers aux étudiants. Rappelons que ces derniers doivent avoir une certaine maîtrise en français oral et écrit pour pouvoir suivre efficacement les cours. Dans ce sens, des cours de techniques d'expression et de communication sont proposés afin de développer chez ces étudiants des compétences de compréhension à la lecture et à l'audition en français.

6 Vers une intercompréhension progressive entre l'amazigh et le français

Aujourd'hui l'intercompréhension est devenue une méthode de communication qui illustre une nouvelle approche dans le champ de la didactique des langues. Dans notre expérience de l'enseignement de l'amazigh à l'Université d'Agadir, nous constatons déjà quelques acquis positifs de cette méthode et son efficacité dans l'intégration de cette langue dans les programmes universitaires. Même si nous ne sommes qu'au début de cette expérience (débutée il y a juste deux ans), nous remarquons qu'une compréhension croisée s'établit petit à petit entre l'amazigh et le français. Cette intercompréhension progressive se manifeste dans plusieurs aspects que nous résumons dans les points suivants :

- rôle majeur de la traduction dans la diffusion des connaissances grammaticales de l'amazigh ;
- valorisation des connaissances linguistiques et du savoir culturel de l'amazigh par l'intermédiaire du français ;
- développement d'un bilinguisme fonctionnel chez les étudiants ;
- émergence d'une situation de coexistence équilibrée entre l'amazigh et le français dans le milieu universitaire ;
- visibilité et place de l'amazigh dans le contexte francophone et dans le champ de la recherche internationale.

Mais, il est trop tôt pour tirer des conclusions ou pour identifier des résultats définitifs de cette expérience. Cela nécessite encore plusieurs années de travail et de réflexion, surtout pour un champ disciplinaire qui est tout à fait nouveau dans le domaine des études amazighes. Nous manquons encore de recherches et d'enquêtes sur le terrain permettant d'examiner les conditions linguistiques et socio-linguistiques qui favorisent ou qui bloquent la communication entre l'amazigh et le français. C'est à partir de ces enquêtes sur le terrain qu'on peut identifier les espaces de communication et d'intercompréhension entre les deux langues. Il s'agit en quelque sorte de reconstituer l'intercompréhension là où la marge de possibilité est plus grande. C'est à partir de cette étape qu'on peut développer une intercompréhension progressive entre les deux langues.

C'est dans ce sens que notre équipe³ envisage d'élaborer quelques perspectives de recherches sur l'intercompréhension. Ces projets s'inscrivent dans le but d'intégrer à moyen terme l'intercompréhension, comme approche didactique, dans le processus actuel de l'enseignement de l'amazigh à l'université marocaine. Il s'agit plus précisément de mettre en place des pratiques communicatives de type intercompréhensif entre l'amazigh et le français.

3 Il s'agit de l'équipe de recherche en didactique de langue et de culture amazighes (ERDLCAM), équipe créée à la Faculté des lettres et des sciences humaines d'Agadir.

Bibliographie

- Andrade (A.) et Moreira (G.), éd., 2002 : *Intercomprehension in Language Teacher Education. Intercompreensão na formação de professores de línguas*, Bilingual Brochure Português/English, Aveiro, University of Aveiro.
- Auroux (S.), 1994 : *La révolution technique de la grammatisation. Introduction à l'histoire des sciences du langage*, Liège, Mardaga.
- Boudreau (A.) et al. éd., 2002 : *L'écologie des langues*, Paris, L'Harmattan (Collection Espaces discursifs).
- El Mountassir (A.), 2009 : *Asselmed n-tchelhit. Méthode de tachelhit (langue berbère du sud du Maroc)*, Paris, L'Asiatheque – Maison des langues du monde.
- Moreau (M.-L.), 1999 : « Pluralité des normes et des appartenances. Convergences et divergences en situation pédagogique », dans *Terminogramme*, n° 91-92, septembre 1999, p. 41-63.
- Puren (C.), 2002 : « Perspectives actionnelles et perspectives culturelles en didactique des langues-cultures: vers une perspective co-actionnelle-co-culturelle », dans *Les langues modernes*, n° 3, 2002, p. 55-76.
- Wambach (M.), 2001 : *Méthodologie des langues en milieu multilingue*, Saint-Ghislain, Centre international audiovisuel d'études et de recherches CIAVER.

*Anna Anastassiadis-Syméonidis,
Maria Mitsiaki,*
Université Aristote de Thessalonique,
Thessalonique,
Grèce,
ansym@lit.auth.gr,
mmitsiak@lit.auth.gr.

*Valentina Chouma,
Eleni Vletsis,*
SNEG,
Thessalonique,
Grèce,
valentinach@yahoo.com,
e_vletsis@yahoo.gr.

Analyse d'erreurs de langue des apprenants du grec langue seconde dans un corpus écrit

Résumé

Le présent article propose de traiter les erreurs dans un corpus de production écrite d'apprenants immigrés. Ce corpus a été le résultat de l'enseignement, au niveau national, du grec langue seconde à des classes de collège pendant deux années scolaires consécutives 2006-2007 et 2007-2008, dans le cadre du projet européen EPEAEK II. 1.1.1.A. Nous avons traité 200 copies venant de 100 apprenants (pour chaque apprenant, une copie datant du début et l'autre de la fin de notre enseignement), dans le but d'en mesurer les résultats. Enfin, nous allons fournir des données statistiques sur le taux d'erreurs phonologiques et morphologiques dans les copies de ces apprenants.

Mots-clés: erreurs phonologiques/morphologiques; corpus écrit d'apprenants; didactique du grec langue seconde.

Introduction

Depuis ces vingt dernières années, la Grèce accueille des immigrés aussi bien des pays balkaniques et des pays de l'ex-URSS que des pays d'Afrique et d'Asie. C'est pourquoi le traitement de l'Autre du point de vue linguistique est un challenge pour le collège en Grèce.

Dans le but d'aider les collégiens étrangers en Grèce dans l'apprentissage du grec moderne, nous avons constitué un corpus de leurs productions écrites comprenant 3 372 copies, qui ont été le résultat d'un enseignement intensif pendant les années scolaires 2006-2007 et 2007-2008. Pour les besoins de cette présentation, nous avons traité 200 copies venant de 100 apprenants. Plus spécialement, nous avons analysé la première et la dernière copie de 100 apprenants, dans le but de mesurer leur progrès en langue, et par conséquent l'effet de notre enseignement pendant deux années scolaires consécutives.

La création d'un corpus à partir de productions écrites d'apprenants étrangers (*learner corpus-based error analysis*, Sinclair 2004) constitue une innovation pédagogique pour l'enseignement de la langue grecque, dans la mesure où elle peut aider, de manière efficace, à la fabrication de matériaux didactiques, à la sélection de documents authentiques, à la constitution d'un syllabus, à l'élaboration de différents types de tests, et susciter le recours aux techniques et aux stratégies d'apprentissage (Meunier 2002 : 122).

Les apprenants dont les copies constituent notre corpus sont du niveau débutant, intermédiaire et avancé. Nous n'avons pris en compte que les erreurs systématiques (Corder 1973). Le modèle de classement des erreurs que nous avons utilisé est fondé sur les niveaux d'analyse linguistique, par exemple, erreurs phonologiques, morphologiques, etc., et cherche à être rigoureux et réutilisable. La comparaison du taux d'erreurs entre les productions des deux années scolaires a révélé une amélioration importante avant tout pour les débutants : a) quantitative, vu la longueur de la deuxième production écrite, et b) qualitative, vu le progrès dans l'acquisition de la structure du système linguistique du grec à tous les niveaux de l'analyse linguistique.

L'analyse des erreurs des productions écrites aide les enseignants à mieux comprendre tant les erreurs dues à la L1 (interférences) que les stratégies utilisées par les apprenants dans le but de surmonter les difficultés. En plus, l'analyse même des erreurs en classe constitue une stratégie d'enseignement cognitive et métacognitive utile et intéressante, puisqu'elle focalise sur l'altérité linguistique.

1 Considérations théoriques

1.1 L'usage linguistique comme source de données

Ces dernières années, il y a une tendance nette à constituer des corpus d'apprenants, puisque, de cette façon, on peut acquérir des productions écrites caractérisées par l'authenticité, la continuité (pas de morceaux choisis arbitrairement) et la variabilité de l'interlangue (le grec comme langue étrangère ou seconde).

Plus particulièrement, il s'agit de données qui fournissent aux chercheurs linguistes et aux enseignants des informations utiles, dans la mesure où elles permettent de voir ce qui est contenu dans la compétence linguistique des apprenants

et de connaître les étapes de leur interlangue. Le champ de recherche de l'acquisition d'une langue comme langue seconde ou comme langue étrangère a pour but d'interpréter la connaissance implicite des apprenants, une connaissance inconsciente qui se manifeste pendant la production spontanée (Ellis et Barkhuizen 2005). Comme nous n'avons pas d'accès direct à la connaissance implicite, nous devons analyser l'interlangue des apprenants telle qu'elle apparaît dans leurs productions écrites et orales. Car, selon Selinker (1972), l'interlangue ne constitue pas une copie imparfaite de la langue cible, mais un système linguistique structuré sur la base de ses propres règles.

Nous considérons que l'analyse de copies d'élèves constitue la façon la plus sûre et valide pour l'étude de l'interlangue des apprenants, dans la mesure où elle fournit des arguments pour la validation ou le rejet des théories d'acquisition d'une langue seconde ou étrangère. Or, pour saisir les diverses étapes de l'interlangue, nous avons besoin de données de bonne qualité (Myles 2005), c.-à-d. de données collectées et organisées selon des critères qui permettent leur mise en valeur.

1.2 Les corpus de copies d'élèves et l'acquisition du grec langue seconde ou étrangère

L'intérêt accru pour le corpus de productions d'apprenants de langue seconde apparaît dans les années 1990 avec le *Longman Learners' Corpus* (Longman Corpus Network 2003) et l'*International Corpus of Learner English* (ICLE, v. Granger 2003). Désormais, il y a une influence mutuelle entre la didactique et la recherche fondée sur le corpus d'apprenants, dans la mesure où ce type de corpus constituerait un miroir de l'interlangue des apprenants. Dans une perspective interventionniste, ce type de corpus permet d'établir un programme dans une position proactive, ainsi qu'un terme de comparaison et un témoignage des étapes franchies par l'apprenant dans une position rétroactive. De même, il serait intéressant de mettre à la disposition des enseignants de grec langue seconde des manuels, des grammaires comparées et des dictionnaires monolingues ou bilingues fondés sur ce type de corpus, qui permet de prévoir les erreurs des apprenants ayant un caractère systématique, particulièrement dans le cas de groupes homogènes.

Le corpus de copies d'élèves utilisé pour ce travail est représentatif et constitue une innovation dans le domaine de la recherche sur l'acquisition du grec langue seconde.

1.3 L'analyse des erreurs dans les corpus de copies d'élèves

Pendant les années 1970, l'analyse des erreurs a été la tendance dominante dans la recherche sur l'acquisition d'une langue seconde/étrangère, avec l'introduction de l'idée que les erreurs peuvent être vues comme des « fenêtres » qui nous permettent de regarder ce qui se passe dans l'interlangue des apprenants et, par conséquent, comme une rétroaction pédagogique utile. Or, cet intérêt pour les erreurs a eu comme résultat la création de typologies d'erreurs fondées sur un nombre important de données d'élèves (Spillner 1991). Dans notre travail, nous procédons à une analyse des erreurs d'apprenants de grec langue seconde, en prenant en compte

des éléments quantitatifs et qualitatifs offerts par cette analyse, dans le but de décrire l'interlangue des apprenants de façon systématique autant que possible, pour apporter de meilleurs résultats dans l'enseignement. Enfin, nous procédons à l'évaluation des apprenants après notre enseignement (voir aussi Taylor et Baker 2008), afin de tirer des conclusions sûres en ce qui concerne l'analyse des erreurs et d'améliorer les méthodes d'enseignement.

2 La constitution du corpus de copies d'élèves

Pour la collecte des données, nous avons établi, avec beaucoup de précautions, les critères pour la constitution du corpus. Nous avons focalisé sur les traits partagés et les traits variables des apprenants ainsi que sur les activités qui nous ont fourni le corpus constitué de productions écrites. Plus spécialement, les apprenants sont tous des étrangers, ils apprennent le grec en Grèce et ils sont inscrits au collège. Toutefois, ils diffèrent sur cinq points : le sexe, les garçons étant plus nombreux que les filles, la L₁, la durée de séjour en Grèce, la durée de scolarisation en Grèce, et, enfin, le niveau de connaissance du grec (*cf.* tableau 1). Nous avons pris soin de noter, pour chaque apprenant, tous les traits précités, pour que notre recherche fournisse des résultats fiables et valides.

En ce qui concerne les textes écrits, ils ont été produits dans les trois cadres communicatifs ci-dessous : description de leur salle de classe à un ami resté dans leur pays d'origine, rédaction d'un message électronique à un ami pour lui exposer leur point de vue sur la façon dont ils collaborent avec leurs camarades de classe, et, enfin, rédaction d'un message électronique à un ami pour lui exposer leur point de vue sur les moyens de transport en Grèce. Comme le genre textuel à produire et la situation de communication étaient les mêmes, les données empiriques étaient comparables.

| Traits communs | | Traits variables | |
|---|-----------------------------|------------------------------------|--|
| Apprenant | Activité | Apprenant | Activité |
| apprentissage du grec langue seconde | Mode de discours (écrit) | Sexe | possibilité d'utiliser des outils de référence (p. ex. des dictionnaires) |
| Age ¹ (apprenants au collège) | Sujet | L1 | |
| | Genre textuel | durée de séjour en Grèce | |
| | | durée de scolarisation en Grèce | |
| | | niveau de connaissances en grec | |

Tableau 1 : Création d'un corpus écrit d'apprenants étrangers : traits communs et variables

3 La méthodologie de la recherche

3.1 Le profil des élèves

Pour la collecte des données de cette recherche, nous avons sélectionné 100 apprenants qui suivaient des cours de grec en plus du programme normal du collège, pendant deux années consécutives (2006-07/2007-08). L'albanais était la L1 de la plupart des apprenants (49), suivi par le géorgien (21) et le russe (11). L'échantillon comprend aussi des apprenants dont la L1 est l'ukrainien, le bulgare, l'anglais, le moldave, etc. (v. Figure 1).

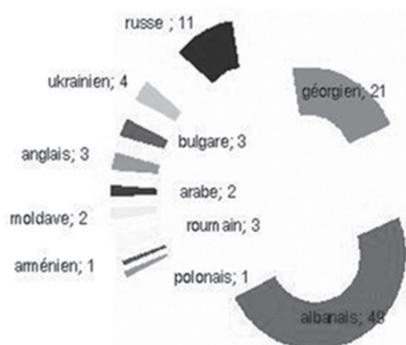


Figure 1: Langue maternelle/langue première

1 Parfois l'affectation en classe ne se fait pas à partir de l'âge des apprenants, mais sur la base de leur niveau de connaissance du grec. Dans notre échantillon, la plupart des apprenants avaient 14-15 ans, l'âge normal pour les élèves au collège.

Un des paramètres pris en compte était la durée de séjour en Grèce (v. Figure 2) : la plupart des apprenants vivent en Grèce plus de 5 ans. Un trait partagé par les apprenants était l'âge : il s'agit de collégiens dont l'âge, à 80 %, se situe entre 13 et 15 ans (v. Figure 3).

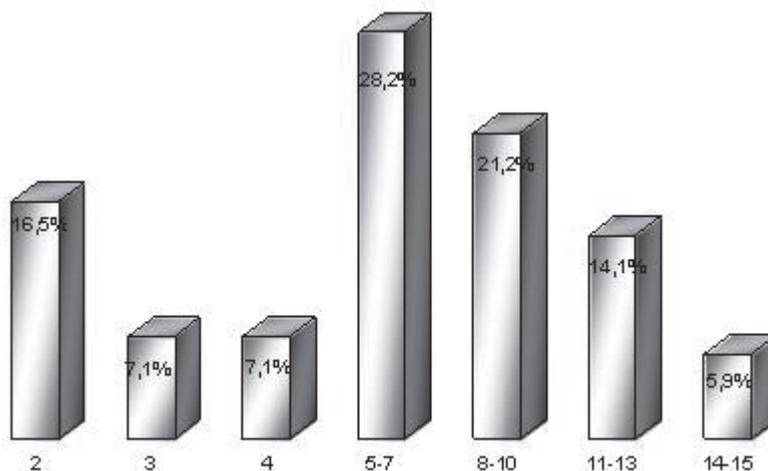


Figure 2 : Durée du séjour en Grèce

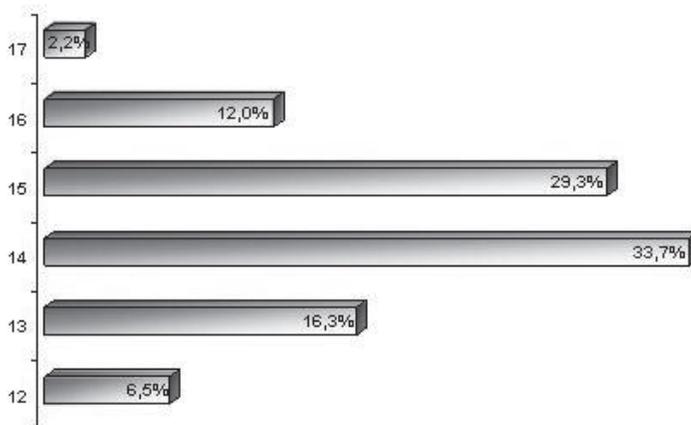


Figure 3 : Âge des apprenants

En ce qui concerne le sexe, les garçons sont plus nombreux que les filles (v. Figure 4).

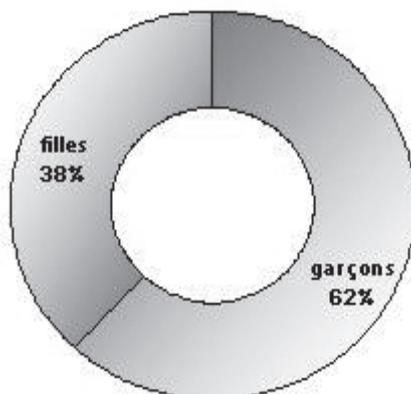


Figure 4 : Sexe des apprenants

3.2 Le modèle de classement des erreurs

L'analyse des erreurs constitue un domaine de la didactologie qui suscite depuis peu l'intérêt des chercheurs (voir Antonopoulou *et al.* 2006, Tzimokas à paraître). Pour apporter plus de rigueur au modèle que nous avons construit, nous avons suivi trois principes : a) le classement des erreurs se base sur le type d'erreur et non sur la cause présumée de son apparition, b) les catégories générales d'erreurs suivent les niveaux de l'analyse linguistique (phonologie, morphologie, syntaxe, sémantique), c) l'absence de chevauchement des catégories créées.

Plus concrètement, les erreurs ont été classées suivant le modèle ci-dessous :

- Phonologie/Mode de transcription : phonèmes, ponctuation, orthographe, accents, mode d'écriture.
- Morphologie : morphème grammatical, morphologie de l'article, genre, aspect, temps, voix.
- Morphosyntaxe : accord en genre, en personne, en nombre, en cas, en mode.
- Syntaxe : omission, addition, emploi erroné (de conjonction, de préposition, d'article, d'argument, de pronom/clitique, de négation), omission de verbe, ordre des mots, structure syntaxique inexistante.
- Sémantique/Lexique : mot inexistant, emploi erroné de mot.
- Marqueurs de l'organisation textuelle : omission, emploi erroné.

4 La typologie des erreurs

4.1 Erreurs de phonologie/mode de transcription

4.1.1 Erreurs de phonèmes

Omission ou addition à l'intérieur des groupes consonantiques, timbre, mode et lieu d'articulation, sonorité.

4.1.2 Erreurs d'orthographe ou de mode d'écriture

Concernant les morphèmes lexicaux et flexionnels, les suffixes, les frontières de mot, l'usage des minuscules à la place des majuscules et vice versa ou le s final, l'emploi de l'écriture phonographique, la présence de lettres de l'alphabet latin², des erreurs de ponctuation et d'accentuation (accents omis ou mal placés).

4.2 Erreurs de morphologie

Concernant la flexion de l'article, le morphème flexionnel, l'aspect verbal, le temps, le genre, la voix.

4.3 Erreurs de morphosyntaxe

Concernant l'accord en genre, en nombre, en cas, en mode, en personne.

4.4 Erreurs de syntaxe

Concernant l'emploi erroné de préposition, de conjonction, l'addition ou l'omission d'article et de pronom, l'emploi de structure syntaxique inexistante, l'omission d'argument, de verbe (surtout de la copule), l'ordre des mots.

4.5 Erreurs de sémantique

Concernant l'emploi erroné de mots ou de mots inexistants.

5 Interprétation

5.1 Phonèmes/Mode de transcription

Le phonème /x/ est problématique chez les albanophones, qui lui substituent /k/³, puisque le /x/ est absent du système phonologique de l'albanais. De même, l'absence du système phonologique de l'albanais et du bulgare de la fricative /ɣ/ conduit les apprenants à la remplacer par l'occlusive /g/. Chez les apprenants russophones l'interdentale /θ/ est substituée par la labiodentale /f/, ce qui est dû à l'absence de l'interdentale dans leur L1. L'omission de la finale *-n* à la troisième personne du pluriel du présent de l'indicatif est due à la forme correspondante du v. de la L1 des

² Pendant les stades initiaux de l'acquisition.

³ Sur 29 productions erronées, cette erreur apparaît 13 fois.

apprenants albanophones, ex. en grec /'exu/ au lieu de /'exun/ (= « ils ont »), car, en albanais, la forme correspondante se termine en une syllabe ouverte, ex. /kanë/.

En ce qui concerne l'orthographe, 60 % des erreurs concernent avant tout les morphèmes lexicaux, mais aussi les morphèmes dérivationnels et flexionnels. La comparaison des deux productions écrites montre qu'il s'agit d'erreurs qui persistent, malgré l'enseignement intensif supplémentaire.

5.2 Morphologie

Indépendamment de la L1 des apprenants, on observe la surgénéralisation, à toutes les classes flexionnelles, du morphème flexionnel *-es* pour la formation du pluriel dans le système nominal. En ce qui concerne l'assignation du genre aux substantifs, on observe la surgénéralisation du neutre, ce qui est dû, d'une part, au fait que la L1 ne possède pas de genre en tant que catégorie grammaticale inhérente au substantif, et, de l'autre, à ce que le neutre constitue pour le grec le genre non marqué (voir Anastassiadis-Syméonidis et Chila-Markopoulou 2003). Un grand nombre d'erreurs concerne l'aspect verbal [continu]- [momentané]⁴. Il est à noter que, dans la seconde production, ce type d'erreur présente une augmentation, ce qui est dû, selon nous, au besoin des apprenants, en passant à un niveau supérieur de connaissances, d'utiliser des structures plus complexes et dans un genre textuel plus exigeant, comme la narration par rapport à la description. De même, on observe chez les albanophones l'emploi du parfait à la place de l'aoriste, comme résultat de transfert dans la L2 de structures de la L1.

5.3 Morphosyntaxe

On observe beaucoup d'erreurs concernant l'accord avant tout en genre, tandis que les erreurs concernant l'accord en nombre se limitent à l'emploi des verbes d'existence. Toutefois, au niveau avancé, il y a une baisse nette du nombre des erreurs concernant l'accord.

5.4 Syntaxe

On observe que la juxtaposition laisse la place à la subordination en passant de la première à la deuxième production écrite. Néanmoins, les erreurs dans l'emploi des conjonctions dans la deuxième production sont dues aux exigences du niveau avancé. Par ailleurs, l'omission de l'article est due à son absence du système du russe et du géorgien. De même, l'addition de l'article défini à des structures avec des verbes d'existence est un résultat de surgénéralisation de son emploi à une place qui aurait dû rester vide.

4 Sur 29 productions erronées, cette erreur apparaît 13 fois.

5.5 Sémantique

On observe des problèmes concernant l'emploi des verbes d'existence, du v. *pijeno* (=aller) avec le sens de « partir », « transporter », et du v. *ksero* (=savoir) avec le sens d'« apprendre ».

6 Remarques conclusives

Il résulte, d'une part, de l'analyse quantitative des données que la plupart des erreurs concernent (i) l'orthographe des morphèmes lexicaux et grammaticaux⁵, (ii) la phonologie et le mode de transcription⁶, (iii) la frontière des mots⁷, (iv) l'accord en genre et en nombre, (v) l'emploi de l'article (omission, addition) et (vi) l'emploi incorrect d'un mot du point de vue sémantique. Voici le nombre d'erreurs par classe :

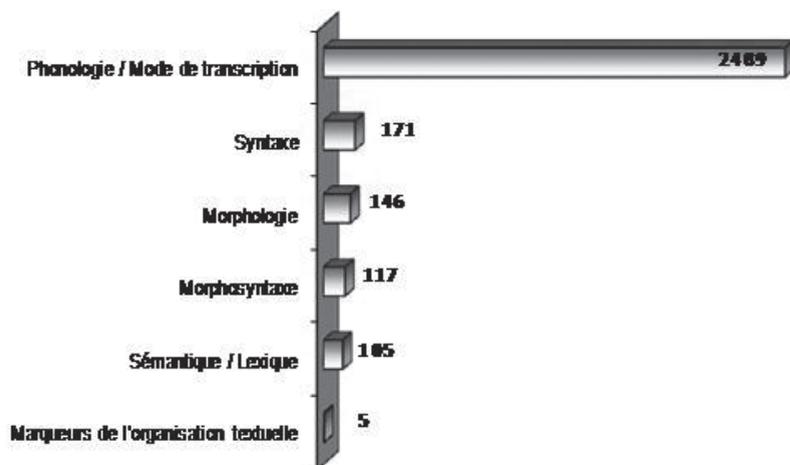


Figure 5 : Présentation quantitative des erreurs

Il résulte, d'autre part, de la comparaison qualitative des deux productions écrites qu'il y a une baisse nette des erreurs morphologiques, avant tout des erreurs qui concernent les morphèmes flexionnels et le genre des substantifs. Néanmoins, les erreurs d'orthographe, avant tout des morphèmes grammaticaux, persistent, ce qui exige, à notre avis, un enseignement explicite et ciblé.

En même temps, on observe une amélioration nette dans l'emploi des structures concernant l'accord en genre et en nombre dans la deuxième production écrite des apprenants, qui représente une étape avancée de leur interlangue. Cette amélioration constitue une preuve de la place centrale de l'accord dans le système du grec.

5 Il s'agit de 60 % de l'ensemble des erreurs, les erreurs de morphèmes lexicaux étant les plus nombreuses.

6 La plupart concernent l'omission de phonèmes, le timbre vocalique, l'écriture phonographique des groupes consonantiques [ks] et [ps] et la suite C [+sonore] + les graphèmes *ia*.

7 Les apprenants, transcrivant le groupe phonologique et pas les mots, séparent le début et la fin du groupe phonologique.

Par ailleurs, les erreurs dues à l'interférence de la L1 sont plus fréquentes parmi les débutants, surtout au niveau phonologique, en comparaison avec les niveaux intermédiaire et avancé. Pourtant, le niveau avancé est à la source de certains types d'erreurs⁸, puisque ce niveau implique l'emploi de la part des apprenants de structures plus complexes et de genres textuels plus exigeants.

Il est à noter que, dans la plupart des copies, l'organisation textuelle est très problématique, vu l'absence des marqueurs de connexité qui caractérisent normalement la structuration textuelle et discursive. Les théories des genres textuels (Swales 1990, Bhatia 1993) peuvent apporter une aide précieuse aux enseignants du grec langue seconde.

Conclusion

Les corpus des copies d'élèves sont un outil extrêmement utile pour la didactique d'une langue seconde, puisqu'ils donnent une image claire des problèmes à traiter et nous incitent à adopter un enseignement focalisé.

Ainsi le processus didactique peut-il être mieux organisé, par le choix a) des faits linguistiques à enseigner suivant les besoins des apprenants, b) de la méthode appropriée suivant le public cible, et c) de matériaux et d'outils adéquats.

Bibliographie

- Anastassiadis-Syméonidis (A.), 2003 : « La didactique de la structure morphologique du grec moderne », *Glossologia* 15, Athènes, p. 25-34.
- Anastassiadis-Syméonidis (A.) et Chila-Markopoulou (D.), 2003 : « Tendances synchroniques et diachroniques du genre en grec - Une proposition théorique », dans Anastassiadis-Syméonidis (A.), Ralli (A.) et Chila-Markopoulou (D.), éd., *Sujets de grammaire du grec moderne - Le genre*, Athènes, Patakis, p. 13-56.
- Anastassiadis-Syméonidis (A.) (en collaboration) : « Questions phonologiques, morphologiques et d'orthographe », *Guide des enseignants*, EPPAS, Thessalonique, 2009, p. 159-176.
- Anastassiadis-Syméonidis (A.) et Mitsiaki (M.) (à paraître) : « Les erreurs phonologiques dans l'interlangue des élèves slavophones du grec langue seconde/étrangère », dans *Actes du congrès international de GALA*, Thessalonique, 2007.
- Anastassiadis-Syméonidis (A.) et Mitsiaki (M.) (à paraître) : « La segmentation morphologique comme stratégie d'enseignement du grec langue seconde/étrangère », dans *Actes du congrès international de GALA*, Thessalonique, 2007.
- Anastassiadis-Syméonidis (A.) (à paraître) : *Sujets de grammaire du grec moderne - L'enseignement du genre*, Athènes, Patakis.

8 Plus particulièrement des erreurs concernant le choix du temps, du mode et de la conjonction introductrice de la subordonnée.

- Anastassiadis-Syméonidis (A.) (en collaboration) (à paraître) : « Erreurs de langue dans un corpus de copies d'élèves de grec langue seconde », dans *Actes du Colloque international sur l'interculturel*, Thessalonique, 12-14 décembre 2008.
- Antonopoulou (N.), Valetopoulos (F.), Karakyrgiou (M.), Moutzi (M.) et Panagiotidou (V.), *Analyse d'erreurs*, www.greek-language.gr/greekLang/files/document/errors/errors_study_a.pdf.
- Bhatia (V.K.), 1993 : *Analysing Genre: Language Use in Professional Settings*, London, Longman.
- Corder (S.P.), 1973 : *Introducing Applied Linguistics*, Harmondsworth, Penguin.
- Ellis (R.) et Barkhuizen (G.), 2005 : *Analysing Learner Language*, Oxford University Press.
- Goutsos (D.), 2006 : « Développement du vocabulaire – Du niveau de base au niveau avancé », dans Goutsos (D.), Sifianou (M.) et Georgakopoulou (A.), *Le grec comme langue étrangère: des mots aux textes*. Athènes, Patakis.
- Granger (S.), 2003 : « The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research », dans *TESOL Quarterly* 37 (3), p. 538-546.
- Meunier (E.), 2002 : « The pedagogical value of native and learner corpora in EFL grammar teaching » dans Granger (S.), Hung (J.) et Tyson (S.), éd., *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*.
- Moser (A.), 1996 : « L'aspect du verbe en grec moderne », dans *Le grec moderne comme langue étrangère: Problèmes d'enseignement*, Athènes, Fondation Goulandris-Horn, p. 77-96.
- Myles (F.), 2005 : « Interlanguage corpora and second language acquisition research », dans *Second Language Research* 21 (4), p. 373-391.
- Papadopoulou (D.), 2005 : « La production de l'aspect verbal d'élèves du grec langue seconde/étrangère », dans *Linguistique appliquée* 21, p. 39-54.
- Selinker (L.), 1972 : « Interlanguage », dans *IRAL*, 10, (3), p. 209-231.
- Sinclair (J.), 2004 : *How to Use Corpora in Language Teaching*, John Benjamins Publishing Co., Philadelphia.
- Spillner (B.), 1991 : *Error analysis: a comprehensive bibliography*, Amsterdam; Philadelphia, J. Benjamins Pub. Co.
- Swales (J.M.), 1990 : *Genre Analysis*, Cambridge: Cambridge University Press.
- Tzimokas (D.) (à paraître) : « Système de classement d'erreurs dans des productions d'apprenants de grec langue seconde », 3^e congrès international sur l'enseignement du grec moderne comme langue étrangère, Université d'Athènes, 22-23 octobre 2004.

Liliana Alic,
Université Transilvania,
Braşov,
Roumanie,
lilianaalic@yahoo.com.

Réussir la traduction

Résumé

Dans cet article, nous nous proposons d'étudier les difficultés qui apparaissent dans l'activité de traduction du français vers le roumain. Le corpus de travail est constitué de publications françaises actuelles et les sujets d'observation sont les étudiants de 3^e année en lettres qui étudient le français comme deuxième langue étrangère. Les difficultés rencontrées par les étudiants sont causées par des analogies avec la langue de base, le roumain, par la connaissance insuffisante de certains aspects de la langue cible – le français –, aspects relevant de la linguistique, de la culture et de la civilisation.

Mots-clés: transfert, équivalent, erreur.

Introduction

Par l'intermédiaire de la traduction, on réalise le transfert d'un énoncé ou d'un texte d'une langue à l'autre. Comme nous ne nous proposons pas de traiter de la traduction littéraire, mais bien de la traduction comme partie intégrante de l'activité didactique, il serait convenable de parler de *transfert linguistique* d'une langue de départ vers une langue d'arrivée, en l'occurrence du français vers le roumain.

La traduction comme méthode d'enseignement d'une langue étrangère est nécessairement doublée d'une analyse contrastive entre les deux langues en contact. Cette analyse contrastive permet à l'apprenant de comparer :

- Les moyens grammaticaux, lexicaux, sémantiques, pragmatiques, stylistiques dont dispose chaque langue pour l'expression de l'information concernée ;
- Les possibilités de transfert d'une langue à l'autre avec le minimum de perte et le maximum de correction.

Ce que nous avons présenté ci-dessus est une situation idéale qui, dans la pratique, est inexistante ou presque. Les difficultés sont diverses et, pour les résoudre, l'enseignant doit travailler à côté de l'apprenant pour trouver des stratégies adaptées à chaque situation. Les deux participants à l'acte de traduction doivent procéder à un inventaire complexe qui vise nécessairement des aspects comme : l'étendue

de l'unité de traduction, le sens du message transmis, les moyens dont l'apprenant dispose en vue du transfert de l'information d'une langue à l'autre, la mise en place d'une stratégie plus ou moins rapide de présentation et d'acquisition de nouveaux moyens, plus appropriés, de réalisation du transfert. Vu qu'il s'agit d'une approche formative de la traduction, il est important pour les deux actants du processus d'enseignement de réfléchir à la mise en place d'une stratégie sûre, accessible et avec des résultats durables visant à corriger les erreurs.

Le contenu du présent article se rapporte à la traduction comme méthode d'apprentissage d'une langue étrangère dans l'enseignement supérieur. Les étudiants sont censés pouvoir transposer un message d'une langue à l'autre, mais ils doivent faire la différence entre les deux systèmes, le système de la langue source et le système de la langue cible. La transposition d'une structure linguistique ou syntaxique d'une langue à l'autre doit se faire conformément aux particularités de chaque langue, et c'est l'enseignant qui doit y veiller.

Le cas particulier de la traduction, du transfert de l'information du français vers le roumain présente une difficulté inattendue : les deux langues sont des langues romanes, donc il y a beaucoup de similitudes, de ressemblances entre les structures de langue. Les apprenants sont alors tentés de tout ramener à un seul système, de négliger les différences entre les deux systèmes, d'emprunter un peu de ceci, un peu de cela à un système ou à un autre sans vérifier si ce qu'ils ont emprunté correspond aux normes du système affecté. Sont-ils conscients que ces interférences peuvent facilement conduire à des fautes interférentielles, parfois difficiles à corriger ?

Dans l'activité de traduction, les étudiants doivent trouver l'équivalent pour un mot, un syntagme, une phrase ou parfois pour une unité plus étendue. Les langues présentent des manières différentes de segmentation de la réalité extralinguistique, de la vie et de l'expérience humaine. Dans ces conditions, une transposition mécanique, automatique des termes d'une langue à l'autre n'est pas une traduction. Il faut trouver l'équivalent le plus approprié qui soit capable de rendre le même sens, la même idée, le même ton, la même intention de communication. Les étudiants rencontrent des difficultés multiples, ayant des causes diverses.

Notre préoccupation va vers les causes qui sont à la base de ces fautes, vers un classement de ces fautes et, pourquoi pas, à l'identification de certaines stratégies de prévention et de correction de ces erreurs de transcodage. Les sujets de cette étude, de cette analyse des erreurs sont les étudiants en 2^e et 3^e années en lettres ou en langues modernes appliquées.

1 Erreurs dues à l'incompétence linguistique

1.1 Erreurs dues à des connaissances insuffisantes de lexique

1.1.1 *Le sens d'un verbe dans une collocation*

Le fait de consulter par un médecin est rendu en français par l'intermédiaire du verbe *passer*: *passer chez le dentiste*. En roumain, *passer chez quelqu'un* signifie lui rendre une visite brève, pour une raison peu importante. Il est dès lors malaisé de comprendre comment Ségolène Royal a réussi à s'offrir un sourire tout neuf en passant simplement chez son dentiste :

[1] « Il y a quelques mois, entre souci esthétique et obligation médicale, *elle est passée chez le dentiste* et s'est offert un sourire tout neuf. » (*Elle*, 23 janvier 2006)

1.1.2 *Le sens d'un syntagme qui contient un numéral*

Pour quiconque a une connaissance moyenne du français, il suffit de savoir que dans une famille, le cadet ou la cadette est le plus petit des enfants, le dernier-né. Vous imaginez le choc de l'apprenant qui se trouve devant le texte suivant qu'il doit traduire :

[2] « Elle assume son bonheur avec Arnaud, *de vingt ans son cadet*. » (*Paris Match*, n° 3102, 2008)

Elle est une anaphore du nom propre *Claire Chazal*.

Une simple information supplémentaire, à savoir qu'« être de vingt ans le cadet de quelqu'un » signifie être plus jeune que cette personne, avoir vingt ans moins qu'elle, suffirait pour clarifier les choses.

Par conséquent, le jeune homme avec lequel Claire Chazal assume son bonheur n'est pas son dernier-né, mais son nouveau compagnon. (Traduction fautive: *Ea își asumă fericirea cu Arnaud, fiul ei cel mic de 20 ani.*)

1.1.3 *Le sens méconnu des néologismes*

Évidemment, les néologismes sont fréquents dans les textes de presse et il n'est pas rare que l'on y trouve des termes empruntés à l'anglais.

[3] « Une candidature Royal serait une mine inépuisable de rebondissements *people*. »

[4] « Ségolène Royal vit actuellement – à son insu ? – le *phénomène* bien connu *de la starisation*. » (*Elle*, janvier 2006)

L'apprenant a beau connaître l'anglais mieux que le français, il se trouve en difficulté dans cette situation précise parce qu'il ne sait pas quel est le sens du mot *people* en français, emprunté et adapté avec le sens de « célébrités, vedettes, stars du cinéma, du sport, de la vie mondaine ». La traduction devra donc incorporer,

d'une manière ou d'une autre, ces informations, ces traits sémiqiques. Pour ce qui est du *phénomène de la starisation*, l'apprenant comprend très bien le sens, mais il est à court de mots pour exprimer ce qu'il comprend : il n'y a pas d'équivalent roumain reconnu et attesté, alors il faut trouver une paraphrase ou une périphrase explicative. La meilleure formulation serait le *phénomène qui consiste à devenir une vedette*.

1.1.4 Traduction des termes inventés

Pour être à même de comprendre un terme inventé, il faut quand même connaître le sens que l'inventeur veut transmettre et aussi connaître le terme qui se trouve à la base de l'invention lexicale.

[5] « Le nouveau métier d'*escort-flic*. Trois sans-papiers sur quatre – 76 % exactement – sont reconduits à l'étranger de façon illégale. [...] En principe, seuls les poulets rattachés à l'Unité nationale d'escorte, de soutien et d'intervention (Unesi) – et pudiquement rebaptisés « escorteurs » – peuvent effectuer les reconduites. » (*Le Canard enchaîné*, 4617, 2009)

Sur le modèle anglais *escort-girl*, le français a trouvé bon d'inventer *escort-flic*, avec le sens de « policier qui escorte les immigrés à la frontière », mais le ton railleur accompagne l'invention lexicale.

Pour la traduction de ce syntagme nominal en roumain, il faut être un peu lexicographe et terminologue. Faute d'hétéronyme, pour la périphrase explicative on pourrait suggérer « policier qui accompagne les immigrés expulsés » « *politist însoțitor al imigranților expulzați* ». Seulement, le ton railleur, moqueur, l'ironie fine ne transparaissent pas.

1.1.5 Erreurs dues à l'ignorance des sigles

Les sigles sont des abréviations plus ou moins récentes de la dénomination d'institutions, d'organismes ou de notions abstraites. Il y a des cas heureux, où les initiales choisies pour le sigle sont identiques dans les deux langues, comme pour *PIB*, mais ces cas sont assez rares. La plupart du temps il s'agit d'initiales distinctes dans les deux langues. Les étudiants ont alors du mal à identifier l'organisme ou toute autre réalité désignée.

En voici quelques exemples :

[6] « À Malte, notre presque président Sarko est sur un bateau : 200 pieds de long (60 mètres), il ne s'agit pas d'un youyou ni d'une périssoire. C'est le yacht de Vincent Bolloré, *un ami du CAC 40* qu'il a rejoint à bord du jet privé du même après une nuit dans la suite présidentielle d'un palace étoilé. » (*Le Canard enchaîné*, mai 2007)

Pour des étudiants roumains, le nom de Vincent Bolloré est tout à fait inconnu et le fait qu'il est *un ami du CAC 40* ne leur dit pas grand-chose non plus. La connaissance du vrai sens de l'abréviation, « Cotation assistée en continu », accompagnée

par la connaissance du fait que cette abréviation désigne un « indice boursier sur la place de Paris employé pour faire un classement des cours des quarante actions les mieux cotées en continu sur le premier marché parmi les cent sociétés dont les échanges sont les plus abondants à l'intérieur de la première Bourse européenne », auraient suffi pour saisir l'implication du syntagme *un ami du CAC 40*. La traduction se fera par une paraphrase (*un familiar al bogaților zilei*) qui est beaucoup plus explicite que le syntagme français.

[7] « En imposant le *CPE*, le premier ministre a pris un grand risque. Il brûle de défendre le *CPE*, mais il lui faut être le plus utile possible et éviter toute déclaration péremptoire. » (*L'Express*, 16.03.2006)

Pour les apprenants roumains, même s'ils ne sont pas débutants, le sigle *CPE* ne revêt aucune signification et ils n'en perçoivent pas les implications. Ils ne savent pas ce que le *Contrat de première embauche* implique et ne comprennent pas les raisons qui ont déterminé les étudiants français à se révolter. C'est ce qui a causé des difficultés dans la traduction du sigle.

[8] « Nicolas Sarkozy fait ses premiers pas dans le cénacle du *G8*. Sur la photo de famille du sommet du *G8*, qui s'ouvre mercredi à Heilingendamm, dans le nord de l'Allemagne, deux dirigeants prendront la pose pour la première fois, deux autres tourneront une dernière fois leur regard vers l'objectif. » (*www.boursier.com*, article du 06.06.2007)

Le *G8* est un organisme international qui prend des décisions importantes pour l'économie des pays les plus industrialisés du monde, mais il faut savoir ce qui se passe dans le monde économique, quels sont les pays qui font partie de ce groupe. En roumain, le sigle est traduit par *Grupul celor 8*.

1.1.6 Les faux amis

[9] « Sur la photo de famille du sommet du *G8*, qui s'ouvre mercredi à Heilingendamm, dans le nord de l'Allemagne, deux dirigeants *prendront la pose* pour la première fois. » (*www.boursier.com*, article du 06.06.2007)

En tant que locuteur roumain on est moins surpris par la traduction du syntagme *prendre la pose* par une paraphrase littérale, ce qui reviendrait à « prendre des photos de quelqu'un », *a-i face cuiva o poză*. Évidemment, il s'agit de deux présidents de deux pays fortement industrialisés qui seront pris en photo dans une posture majestueuse, conforme à leur statut.

2 Erreurs de syntaxe

2.1 Difficulté à identifier le sujet de la phrase

La présence du sujet ou du pronom sujet est obligatoire dans toute phrase française. Mais que faire quand on se trouve devant deux pronoms dont on ne sait pas lequel est le pronom sujet ?

- [10] « *Il nous faut* mettre un terme au scandale que représente la non-représentation d'un grand parti comme le Front national à l'Assemblée nationale », répète-t-il à chaque occasion. (*www.lepoint.fr*, 01.06.2007)

La traduction fautive est : *Ne trebuie un termen pentru scandal*, faute qui a été engendrée par le mauvais choix du pronom sujet et qui a conduit vers une traduction totalement erronée.

La difficulté aurait pu être évitée si l'apprenant s'était fait une réflexion supplémentaire : le sujet d'un verbe impersonnel comme *falloir* est *il*, singulier.

2.2 Difficulté à identifier une relation logique

Les relations logiques constituent un problème dans le passage d'une langue à l'autre, surtout quand elle ne revêt pas sa forme canonique. La relation logique de concession réalisée par l'intermédiaire de deux propositions indépendantes n'est pas si facilement identifiable pour les apprenants.

- [11] « De cette insoumission, elle gardera l'envie de prouver qu'elle mérite mieux que la voie qu'on lui a tracée. *Quitte à devenir à son tour dure et autoritaire.* » (*Elle*, 23 janvier 2006)

2.3 Polysémie d'un relateur

Un *si* dans une phrase oriente d'emblée l'apprenant vers la relation logique de condition. Mais les relations logiques réalisées par ce relateur sont multiples, il faut convoquer des connaissances plus poussées pour pouvoir traduire correctement la phrase suivante, où la concession est difficilement saisissable pour l'apprenant :

- [12] « *Si* elle permet à Jean-Marie Le Pen de sillonner la France « en un temps record », cette stratégie ne lui laisse guère le loisir de rencontrer les électeurs. » (*www.lepoint.fr*, 01.06.2007)

3 Erreurs dues à la connaissance insuffisante de la morphologie

3.1 L'infinitif de narration

Selon les règles de la syntaxe, un verbe à l'infinitif n'est pas un réalisateur du prédicat. À moins que ce ne soit un infinitif de narration. Il s'agit là de la connaissance insuffisante des valeurs de l'infinitif. Dans un texte narratif, l'infinitif peut jouer le rôle de prédicat, même sous cette forme dépourvue de marques de temps et de personne.

Dans l'article « Tous coluchiens, même Giscard » on parle de l'incompatibilité entre l'artiste et le politicien, les deux ayant critiqué d'une manière plus ou moins élégante l'autre. À côté des difficultés causées par l'ignorance du contexte social, la traduction est rendue encore plus difficile par la présence de cet infinitif de la

narration. Les apprenants ont du mal à traduire une phrase qu'ils considèrent dépourvue de prédicat :

- [13] « Coluche giscardien ! Le secret, jusque-là, était bien gardé. L'ex l'a révélé lundi, sur Europe 1 : « Je l'appréciais beaucoup et je savais qu'il était pour moi, en fait. » [...] Et Giscard *de décrire* son Coluche à lui. » (*Le Canard enchaîné*, n° 4590, 2008)

3.2 Le rôle d'un pronom adverbial

Le pronom adverbial est un substitut. Il suffit de savoir ce qu'il remplace pour réussir à comprendre une phrase.

- [14] « Un chiffre impressionnant : 94 % des Français y sont favorables. » (Elle, janvier 2006)

L'apprenant chargé de traduire cette phrase dans le contexte d'un vote favorable pour Ségolène Royal de la part d'un grand nombre de Français ne savait pas que le pronom adverbial remplace un complément exprimé par un substantif précédé de la préposition à : « être favorable à quelque chose ». Le contexte le plus connu dans lequel un apprenant rencontre le pronom adverbial y est comme substitut d'un complément de lieu : « aller au marché – y aller ». Ne connaissant pas l'autre contexte d'emploi de y, l'étudiant roumain l'a simplement ignoré dans la traduction.

3.3 Valeurs d'un participe passé

En français comme en roumain, le participe passé peut avoir une valeur adjectivale, ce qui le rend apte à la nominalisation : *les morts*, *les licenciés*, *les interviewés*, *les disparus*. Dans une traduction du français vers le roumain, nous devons faire ressortir la valeur adjectivale du participe passé, en ajoutant ceux devant certaines catégories obtenues par conversion, sans quoi le sens est ambigu ou la traduction fautive.

- [15] « De plus, depuis quelques années, on sait à quel point experts et sondages se trompent. Comme si *les sondés* disaient, dans les interviews, autre chose que ce qu'ils exprimaient dans les urnes. » (*Elle*, janvier 2006)

4 Erreurs dues à l'incompétence sémantique

4.1 La polysémie

Devant un terme polysémique, on peut avoir des hésitations, mais les apprenants peuvent se méprendre sur le sens d'un substantif commun très fréquent.

- [16] « Des enseignants m'ont dit que des petits garçons leur ont demandé si un homme pouvait être président. » (*Elle*, janvier 2006)

L'apprenant roumain va traduire *homme* par « être humain » (en roumain *om*, opposé à « animal »). En réalité, la question des enfants de l'école maternelle de

France a été si un « être humain, adulte, mâle » (opposé à « être humain, adulte, femelle ») peut remplir la fonction de président de la République. Les enfants, dépourvus d'expérience en matière de politique, ne connaissaient que M^{me} Royal comme candidat à la présidence.

4.2 Le sens des expressions plus ou moins figées et des proverbes

[17] « De plus, depuis quelques années, on sait à quel point experts et sondages se trompent. Comme si les sondés disaient, dans les interviews, autre chose que ce qu'ils exprimaient dans les urnes. Ainsi, en 2002, personne *n'a vu venir* Le Pen au premier tour, pas plus qu'en 2005 le vote protestataire du « non » à l'Europe. » (*Elle*, janvier 2006)

Voir venir quelqu'un est une expression dite idiomatique qui signifie « prévoir l'arrivée de quelqu'un, son immixtion dans une affaire, prévoir ses intentions ». Elle ne signifie nullement « regarder quelqu'un venir, s'approcher », mais faute de connaître le vrai sens de l'expression idiomatique, on se tourne vers le sens littéral qui n'est pas le sens exact dans ce contexte.

[18] « Souvent, les femmes politiques s'autolimitent, par doute ou manque de confiance en elles ; elles ont intériorisé malgré elles les stéréotypes sur les sexes. Ségolène a dépassé ce stade. Elle est sortie du bois toute seule, avec cran. » (Citation de Christophe André, psychiatre)

L'expression idiomatique « sortir du bois » signifie « dévoiler ses intentions ». Il ne s'agit pas du tout du sens littéral qui pourrait exister et qui pourrait être pris en compte dans un tout autre contexte, par exemple un contexte situationnel.

L'un des éditoriaux de Claude Imbert est intitulé « L'hirondelle » (*Le Point*, mars 2007) et il parle évidemment de Ségolène Royal, en faisant allusion au proverbe *Une hirondelle ne fait pas le printemps*. Le sens de l'article est difficile à saisir à cause du caractère métaphorique qui prévaut, mais également à cause du caractère idiomatique des deux langues. Chacune d'elle a choisi un autre élément de la nature qui annonce le printemps : le français a choisi l'hirondelle, d'où le proverbe *Une hirondelle ne fait pas le printemps*, tandis que le roumain a choisi la fleur, génériquement, en concevant le proverbe *Cu o floare nu se face primăvară*.

4.3 L'identification des changements de sens métagénériques (la métonymie)

[19] « De plus, depuis quelques années, on sait à quel point experts et sondages se trompent. Comme si les sondés disaient, dans les interviews, autre chose que ce qu'ils exprimaient dans les urnes. » (*Elle*, janvier 2006)

Urne est un terme qui se trouve dans une contiguïté causale avec les élections, donc l'expression est une métonymie qui signifie « voter pour quelqu'un ou pour quelque chose ». Il y a même une métonymie semblable en roumain *a se duce la urne* (« voter »). Par conséquent, la traduction ne devrait pas être difficile, mais le

changement de sens intervenu sème le doute chez l'apprenant, qui se trouve devant le dilemme : les choses sont-elles identiques dans les deux langues ?

[20] « Ségolène Royal n'a pas trouvé indécent d'*aérer ses escarpins* dans les quartiers pauvres de Puente Alto. » (*Elle*, janvier 2006)

Aérer ses escarpins dans un quartier cela signifie « se promener dans ce quartier », mais cela s'exprime toujours par métonymie (contiguïté causale). La contiguïté reste à être identifiée par l'apprenant, qui hésite devant cette manière de s'exprimer du journaliste.

[21] « Christine Lagarde voit des paresseux partout ! Dans son envolée à L'Assemblée, la ministre a décrété : "Il faut cesser de penser et *se retrousser les manches*". » (*Le Canard enchaîné*, n° 4525, 2007)

Combien d'apprenants pourraient identifier le vrai sens de la métonymie et pourraient la traduire par un équivalent de « se mettre au travail » ?

4.4 Connotation manquée

La plupart du temps, la connotation échappe à l'apprenant d'une langue étrangère, car il n'est pas au courant des événements sociaux, économiques, politiques ou autres.

[22] « Chirac aimait les suites au Royal Palm, et observait les yachts à la jumelle. En partant, il va se faire héberger par la famille de son défunt et riche ami Hariri. Sarkozy, lui, c'est Bolloré, il arrive en yacht de 200 pieds, il pratique la rupture en *Falcon* et au *Fouquet's*. » (*Le Canard enchaîné*, mai 2007)

Beaucoup de termes de la citation ont, en plus de leurs sens dénotatif, un sens supplémentaire, connotatif, connu de quelques initiés ou des membres d'une communauté.

Comme les apprenants roumains ne sont pas au courant des fréquentations de l'ancien président ou de celles du président français en fonction, ils ne peuvent pas faire de connexions et les sens ajoutés, les connotations, se perdent. Il en va de même pour les deux noms propres, *Falcon* et *Fouquet's*, qui représentent l'avion des riches et le restaurant chic situé sur les Champs-Élysées, où le président gagnant a fêté sa victoire.

5 Erreurs dues à l'incompétence textuelle

La valeur de substitut du pronom *ce*

[23] « Elle est arrivée, la voilà. 59 % des Français voteraient – « certainement » ou « peut-être » – pour Ségolène Royal. *Et ce*, en dépit des déclarations misogynes au sein de son propre parti. » (*Elle*, janvier 2006)

Ce est dans ce cas un substitut pour toute une phrase : « 59 % des Français voteraient — « certainement » ou « peut-être » — pour Ségolène Royal. » Faute de connaître la valeur de cet humble substitut, on a du mal à traduire le paragraphe. Et du point de vue de la linguistique textuelle, le pronom neutre *ce* a une valeur anaphorique, ce qui est assez rare pour ce pronom. Il figure dans beaucoup de contextes avec sa valeur cataphorique (p. ex. : *Ce qui m'intéresse c'est la littérature*, avec sa valeur anticipative) et comme cette valeur anaphorique est moins employée dans des textes à usage didactique en vue de la traduction, elle reste méconnue.

6 Erreurs dues à des connaissances insuffisantes concernant l'histoire de France

[24] « La société française traverse une période de grand trouble. Toutes ses incertitudes, tous ses repères, toutes ses traditions, toutes ses grilles d'interprétation, tous ses comportements classiques vacillent, hésitent et tremblent. Depuis la fin de la Seconde Guerre mondiale, elle s'était habituée à une croissance ininterrompue : avant la crise, elle a réappris l'inquiétude, l'improvisation et l'instabilité. Pendant les *Trente Glorieuses*, le plein emploi allait de soi. » (« Regard sur la société française », M. Bardin)

Traduire les trente glorieuses par « l'époque de trente ans après la guerre (en réalité 28 ans) qui représente une période de forte croissance économique qu'ont connu entre 1945 et 1974 une grande majorité des pays développés, principalement les membres de l'OCDE » c'est un peu long, mais c'est correct. Faire une paraphrase littérale et traduire mot pour mot, cela n'est pas correct, mais c'est ce qui se passe si les apprenants ne connaissent pas la vraie signification du syntagme. Le résultat est *In timpul celor treizeci de glorioase*, syntagme qui n'a pas de sens en roumain.

7 Erreurs dues à la connaissance insuffisante de la vie socioculturelle française

L'article « Tous coluchiens même Giscard » contient une reprise des disputes et des coups fourrés échangés entre Giscard d'Estaing et Coluche.

[25] « *Coluche giscardien* ! Le secret, jusque-là, était bien gardé. L'ex l'a révélé lundi sur Europe 1 : » Je l'appréciais beaucoup et je savais qu'il était pour moi, en fait. Il n'était pas du tout pour François Mitterrand. » » (*Le Canard enchaîné*, n° 4590, 2008)

Quelques connaissances élémentaires de la couleur politique des présidents français sont indispensables pour pouvoir comprendre le texte correctement, c'est-à-dire en saisir l'ironie. De plus, la couleur politique de Coluche est bien connue pour ceux qui connaissent l'histoire de sa vie. Pour les apprenants du français langue étrangère, les difficultés commencent au moment où on leur demande d'expliquer l'ironie de la deuxième phrase.

8 Erreurs dues à la méconnaissance du niveau de langue

La traduction est une opération de transfert d'informations d'une langue à une autre. Elle se fait d'une manière très correcte, ou enfin, elle devrait se faire ainsi, quand il s'agit du niveau standard de la langue et de traduction du degré zéro d'écriture auquel faisait allusion Roland Barthes.

Dès qu'il s'agit de dépasser ces limites, dès que le niveau de langue n'est plus le niveau standard, mais qu'il glisse légèrement vers le langage familier, le sens échappe aux apprenants, même à ceux qui ont un niveau moyen de connaissances.

[26] « 22, voilà Vingt-Trois! Jeudi dernier, 6 novembre, le cardinal-archevêque de Paris, André Vingt-Trois, était, de 17 à 18 heures l'invité de « Face aux chrétiens », une émission conjointement organisée par le quotidien « La Croix, RCF et Radio Notre-Dame. » La discussion tournait autour de la question de savoir si les femmes peuvent lire l'Évangile pendant la messe. La réponse de l'archevêque a été qu'il n'y voyait pas d'inconvénient, pourvu que les femmes soient formées, et que « l'important ce n'est pas d'avoir une jupe, mais d'avoir quelque chose dans la tête ». (*Le Canard enchaîné*, n° 4594, 2008)

Il paraît que la signification de l'exclamation *Vingt-deux!* en argot est inconnue des étudiants roumains qui n'ont eu que peu ou pas de contact avec ce niveau de langue. Ils se trouvent dès lors dans l'impossibilité de saisir l'ironie du jeu de mots « 22, voilà Vingt-Trois! », basé sur l'exclamation argotique et le nom du cardinal-archevêque de Paris, *Vingt-Trois*, qui semble avoir des idées et des conceptions très « XXI^e siècle ».

Conclusion

L'activité de traduction suppose, outre une bonne connaissance des deux langues en contact, des connaissances ayant trait à leur culture et à leur civilisation, le tout accompagné d'une connaissance minimale des réalités sociales, économiques et culturelles actualisées des pays concernés, en l'occurrence la France et la Roumanie. Les équivalents donnés pour les termes français qui n'ont pas de correspondant roumain doivent être en concordance avec l'esprit de la langue roumaine et, s'il n'y a pas d'attestation officielle du terme en roumain, la solution la plus correcte est de donner une périphrase explicative.

Bibliographie

- Ballard (M.) et El Kaladi (A.), 2003 : *Traductologie linguistique et traduction*, Artois Presses Université.
- Cristea (T.), 1977 : *Éléments de grammaire contrastive. Domaine français-roumain*, București, Editura Didactică și Pedagogică.
- Cristea (T.), 1998 : *Stratégies de la traduction*, București, Editura Fundației România de Măine.

Gherasim (P.), 2003 : « Expression linguistique de la subjectivité dans le discours et le discours rapporté », dans Moeschler (J.), dir., *Cahiers de linguistique française*, n° 25, p. 205-220, Genève, Université de Genève.

Marquilló Larruy (M.), 2003 : *L'Interprétation de l'erreur*, Édition Christine Grall, CLE International/VUEF.

Lamria Chetouani,
IUFM de Bretagne/Université Bretagne Occidentale – CREAD,
Rennes,
France,
lamria.chetouani@bretagne.iufm.fr.

Le français transversal vecteur des enseignements/apprentissages des disciplines scientifiques

Résumé

L'étude porte sur le vocabulaire commun qui traverse les disciplines scolaires et qui se situe à l'interface de plusieurs domaines de connaissance. Le traitement quantitatif et qualitatif du champ d'investigation oral et écrit met en évidence un noyau lexical invariant indispensable pour la structuration, la cohérence et le fonctionnement du discours scientifique tout en incarnant une part non négligeable de sa surface textuelle. Dans la perspective d'une meilleure réussite scolaire, l'appropriation des savoirs scientifiques scolaires requiert une connaissance suffisante de ce français accompagnateur des termes techniques.

Mots-clés : vocabulaire ; scientifique ; non terminologique ; transdisciplinaire.

Introduction

Cette contribution a pour objet l'étude de la langue scientifique, et plus particulièrement du vocabulaire transdisciplinaire qui transcende les domaines de spécialité et qui s'inscrit dans une dimension linguistique globale, tout en côtoyant les termes spécialisés inhérents à tel ou tel domaine particulier. Il relève à la fois de la langue courante et de la langue scientifique dont il fait partie intégrante et dont il est indispensable non seulement pour la structuration du discours, mais aussi pour l'argumentation, l'explication, la transmission didactique des savoirs scolaires, voire pour la communication scientifique en général.

L'intérêt de l'étude est double, didactique et scientifique. D'une part, ce vocabulaire participe à la construction du savoir scientifique et contribue à son appropriation, tout en alimentant la question de transversalité inhérente à la nouvelle philosophie éducative d'aujourd'hui. D'autre part, il complète les études lexicales sur les langues spécialisées. La nécessité de le décrire et de le faire connaître est motivée par la question de la transdisciplinarité, devenue aujourd'hui une exigence primordiale du *Socle commun de connaissances et de compétences* (MEN 2006) lequel incarne le levier de réduction des inégalités sociales et de réussite scolaire. L'étude sur le *Vocabulaire général d'enseignement scientifique* (Chetouani, 1997)

met en évidence le langage vecteur des apprentissages scientifiques, par opposition aux termes spécialisés. Son titre, bien qu'évoquant le livre d'André Phal (1971) s'oppose fondamentalement à lui, car le *Vocabulaire général d'orientation scientifique* (VGOS) se veut résolument scientifique, caractérisant « la substantifique moelle » des sciences dites « dures ». Le français transversal, dont il est question ici, représente, au contraire, un français omniprésent dans tout discours spécialisé dont il constitue la trame linguistique ; il correspond à la compétence langagière servant à relier les savoirs et les savoir-faire. À ce titre, il se distingue, à la fois, du concept traditionnel de « langue de spécialité » qui concerne toutes les composantes linguistiques (syntaxe, morphologie, sémiologie, etc.) du discours technoscientifique (Kocourek, 1982 : 33-85), et de la terminologie proprement dite, qui tend à isoler les spécialités. Car, comme le souligne P. Lerat (1995 : 21), « une langue spécialisée ne se réduit pas à une terminologie, elle mobilise 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 » : les termes techniques s'articulent sur ce fond commun en langue naturelle.

Parmi les recherches récentes sur le français scientifique, celle de Coxhead (2000) concernant le noyau lexical académique des disciplines scientifiques se rapproche par sa méthode et par sa visée didactique de notre étude, mais elle concerne l'anglais langue seconde et s'inscrit dans le sillage du *Basic English* d'Ogden.

D'autres recherches intéressantes, mais n'ayant rien à voir avec le vocabulaire de base dont il est question ici, s'occupent du discours scientifique du point de vue énonciatif, pragmatique, rhétorique, terminologique ou phraséologique (Tutin 2007b).

Le lexique transversal proprement dit a fait l'objet de la recherche de P. Drouin (2007) qui sélectionne les mots au moyen des spécificités positives (logiciel de P. Lafon) couplées avec le critère de répartition et qui s'intéresse aussi aux collocations lexicales tout comme celle d'A. Tutin (2007a). Celle-ci, qui procède au moyen de méthodes distributionnelles de traitement automatique des langues (TAL), s'attache particulièrement au traitement des substantifs, à travers leurs relations de sous-catégorisation et leurs classements sémantiques. Ces différentes recherches suivant des approches spécifiques sont complémentaires et contribuent à une meilleure appréhension du langage scientifique.

Notre approche méthodologique, à partir d'un corpus non exhaustif, se distingue par son esprit et sa visée didactique dans la lignée des travaux du CREDIF ; en tant que « vocabulaire de base » la liste de mots envisagée se veut au service d'une acquisition propédeutique des disciplines scolaires.

L'aspect linguistique de ce vocabulaire transversal (VT) sera abordé de deux points de vue complémentaires : discours de professeurs et productions d'élèves. Dans un premier temps, à partir de l'analyse d'un corpus pluridisciplinaire formé de cours produits par des professeurs de lycée, sera mis en évidence un ensemble de mots caractérisant divers enseignements scientifiques. Dans un deuxième temps,

la description des difficultés linguistiques des élèves du primaire, ciblées sur les obstacles du français langue maternelle lors de la résolution de problèmes mathématiques, débouchera, dans la troisième partie, sur des réflexions pédagogiques et didactiques à propos de la transdisciplinarité.

1 Le français transversal dans les pratiques enseignantes

Des analyses ont été effectuées à partir d'un corpus de cours en sciences dites « dures » et en sciences sociales et humaines dans des établissements secondaires algériens où le français a un statut de « langue d'enseignement scientifique » ; cet enseignement est conforme au programme en vigueur en France. Le champ d'investigation est constitué de textes extraits de manuels scolaires et, majoritairement, d'enregistrements de cours oraux produits par des professeurs lors de communications réelles en situation de travail authentique. Le corpus est certes réduit vu la rigueur qui s'impose dans le comptage manuel et l'exploitation fastidieuse de l'oral (Chetouani 1997 : 39-50), mais il prétend être représentatif du discours scientifique scolaire dans la mesure où la variation des pratiques a été nécessaire pour exploiter plusieurs situations de productions langagières : niveaux scolaires différents (de la seconde à la terminale), formes de cours variées (TP, cours, exercices, composition, correction, cours éparpillés, cours regroupés), variété de disciplines (mathématiques, physique-chimie, droit et économie) et diversification de contenus (par exemple, pour la physique : électricité, dynamique, statique, chimie).

1.1 La méthode d'analyse

Après un travail préliminaire d'enregistrement et de transcription du corpus, le comptage des mots a été mené suivant l'option de lemmatisation et d'exclusion des chiffres, signes et symbolisations scientifiques. Des classements et des tris ont été effectués à partir de critères grammaticaux et statistiques (fréquence, répartition, utilisation). Quatre listes de référence ont servi d'appui pour l'étude comparative et la recherche des convergences : le *Français fondamental* (FF) de Georges Gougenheim *et al.*, le *Vocabulaire général d'orientation scientifique* (VGOS) d'André Phal, le *dictionnaire de fréquences du Trésor de la langue française* (TLF) de Robert Martin *et al.*, et le *Frequency Dictionary of French Words* (FDFW) d'Alphonse Juilland *et al.*

Cette série de critères quantitatifs a permis de dégager le vocabulaire transversal selon deux procédés lexicologiques privilégiés : l'intersection des vocabulaires de corpus hétérogènes et, comme critère de différenciation, le fait que ces corpus soient marqués positivement ou négativement en termes de communication scientifique. Ainsi, sont écartés de la liste : les mots n'apparaissant que dans la liste de sciences exactes (c'est-à-dire les termes spécifiques), les mots du VT qui n'ont de correspondants que dans la liste du VGOS (donc des mots scientifiques) et, enfin, les mots du VT attestés dans TLF et/ou FDFW, avec des divergences de fréquences.

Le croisement de ces procédures comparatives a donné une liste de mots classés alphabétiquement et par ordre décroissant de fréquences avec des indications grammaticales, fréquentielles et pédagogiques (appartenance au FF et au VGOS). Le traitement manuel a permis aussi de faire un sort à tous les mots polysémiques. Ainsi, par exemple : le verbe « (se) trouver » (une entrée), mais 2 ou 3 formes pour : « que » (relat, conj. sub, et inter), « fin » (N et adj), « comme » (compar., cause), « autre » (N, adj)..., et 5 formes pour le « l' » (art fem., art masc., p. pers fem., p. pers masc., forme superfétatoire de « l'on »).

1.2 Description de la liste du VT

1.2.1 Du point de vue numérique

La liste se compose de 832 unités lexicales, soit plus de la moitié du nombre total de mots du corpus, qui compte 1 424 mots lemmatisés et 26 026 occurrences. Elle est relativement restreinte par rapport à d'autres listes de vocabulaire, comme celles du FF (1 060 mots), du VGOS (1 160 mots), de Haygood (2 069 mots), de Henmon (3 905 mots), de Juilland (5 082 mots), de Vander Beke (6 067 mots).

Cependant, un vocabulaire dit « de base » est nécessairement limité : plus la liste est courte, plus elle est, didactiquement parlant, utile et plus elle est longue, plus elle est difficile à didactiser. La liste de Juilland, par exemple, qui a une visée davantage culturelle que pédagogique, définit préalablement le nombre de ses unités (5 000 environ) pour chacun des dictionnaires relevant des cinq langues romanes. Elle ne prétend donc pas recenser seulement les mots les plus « utiles ».

La liste VT est obtenue au moyen des critères de fréquence, répartition et « utilisation » (respectivement : Fq, Rep et Ut).

En termes de pourcentages, le contenu du VT se traduit par : 316 mots (38 % du total) fournis par le critère de convergence de Fréquences, 214 mots (25,75 %) produits par le critère de Répartition (sciences exactes + sciences humaines), et 302 mots (36,25 %) résultants du critère d'Utilisation (fréquence x répartition).

Le chevauchement des trois critères recouvre 33,4 % du total ($\frac{1}{3}$ du contenu de la liste). Aussi, le VT est constitué de : 233 mots communs aux deux listes pédagogiques (FF et VGOS), 138 mots appartenant seulement au FF, 227 mots appartenant seulement au VGOS, et 234 mots « originaux », c'est-à-dire n'appartenant ni au FF ni au VGOS.

1.2.2 Du point de vue grammatical

On peut compter 292 noms (34,15 %), 283 verbes (29,46 %), 110 adjectifs (11,98 %) et 197 mots grammaticaux (24,41 %). Ces pourcentages sont très intéressants.

1. Les noms dont la présence est caractérisée par le taux le plus élevé des quatre catégories grammaticales représentent plus d'un tiers du total. Ce phénomène est

remarquable en comparaison avec d'autres listes. L'indigence en substantifs du *Français fondamental* (FF) a conduit au recours aux critères de « disponibilité » et de « centres d'intérêt », qui n'ont pas échappé à de violentes critiques.

2. Les verbes, presque aussi nombreux que les noms, prouvent que la tendance à la réduction des phrases (par ellipse, nominalisation, adjectivation...), chère à la langue scientifique, ne semble pas s'appliquer à la langue transversale, qui fonctionne avec beaucoup de verbes.

3. Les mots grammaticaux, dont le nombre est un peu inférieur au quart du total, semblent être indispensables à la structuration et à l'intelligibilité du discours (déterminants, prépositions, connecteurs temporels, etc.). Ce type de mots est exclu des listes de vocabulaire scientifique telles que le VGOS ou les terminologies, et même du FF qui élimine les têtes de liste représentées par les mots les plus fréquents.

4. Les adjectifs enfin, sont les moins nombreux. Là aussi, le vocabulaire scientifique se distingue du vocabulaire transversal, en ce sens qu'il a tendance à employer l'adjectivation comme moyen de condensation des phrases (Kocourek 1982 : 68).

En somme, bien que les adjectifs soient peu nombreux comparativement aux autres parties du discours, le vocabulaire dénotatif formé de noms et d'adjectifs (402 mots, soit 48,3 % du total) est presque dans les mêmes proportions que le vocabulaire fonctionnel, formé de verbes et de mots grammaticaux (430 mots = 53,87 %). Ce dernier, légèrement supérieur au premier, contient toutes sortes de verbes : verbes d'état (*être, paraître, sembler...*), verbes d'action et de manipulation (*faire, observer...*), verbes concrets (*tracer, souligner...*) et abstraits (*déduire, concevoir...*), verbes de parole (*dire, définir...*), etc. Il contient également toutes sortes de mots servant de liens intra- ou interphrastiques, voire intersémiotiques. Les liens entre les phrases et dans les phrases sont des outils d'expression et de pensée qui peuvent être groupés par catégories grammaticales (conjonctions de coordination, de subordination, prépositions, adverbes, pronoms, déterminants...) ou par catégories fonctionnelles. Dans ce champ, par exemple, sont attestés divers connecteurs : additifs (*et, encore, de plus, aussi, également...*), illustratifs (*par exemple, en particulier, notamment...*), énumératifs (*d'abord, ensuite, enfin, finalement, premièrement...*), explicatifs (*car, c'est-à-dire, aussi, autrement...*), comparatifs (*aussi, ainsi, plutôt, mieux, plus, moins...*), adversatifs (*mais, or, en revanche, au contraire, cependant...*), concessifs (*toutefois, cependant, néanmoins...*), temporels (*d'abord, ensuite, puis, alors, en même temps...*), causaux/consécutifs (*donc, car, ainsi, c'est pourquoi, en conséquence...*), etc.

1.2.3. Du point de vue thématique

Les mots peuvent être groupés grossièrement en quatre domaines principaux.

1. Le domaine énonciatif et communicationnel est caractérisé par des déictiques personnels (*je/moi, nous, vous*), des embrayeurs spatio-temporels (*hors, place,*

année, aujourd'hui...), des termes appréciatifs ou évaluatifs et des marques subjectives : adverbess d'opinion, adjectifs appréciatifs, modalisateurs, etc. (bien, assez, juste, faux, exact, très, important, lentement, littéralement, nettement, tout à fait, pouvoir, devoir...).

2. Le domaine pédagogique concerne le comportement enseignant et la pratique pédagogique : métalangage, verbes de parole (*définir, énoncer, expliquer, répondre, interroger, récapituler, résumer...*) et mots renvoyant aux activités de classe, physiques et mentales (*application, comprendre, constatation, montrer, écrire, correction, utiliser, produire, démontrer, prouver, rechercher, représenter, simplifier...*).

3. Le domaine scientifique s'exprime par référence au domaine empirique : termes relatifs aux manipulations et aux expériences (*accroître, augmenter, courber, schématiser, mesurer, machine, électricité, construire...*) ou par rapport au domaine intellectuel qui relève soit des calculs ou des notions élémentaires de mathématiques (*addition, multiplication, diviser, résoudre, retrancher, résultat, somme, égal...*), soit du raisonnement logique (*déduire, réciproque, parce que, donc, étant donné...*) et de la pensée scientifique (*hypothèse, si, mettons...*).

4. Le domaine institutionnel ou socioculturel se rapporte, enfin, au cadre scolaire (*cours, matière, liste, ordre, silence, ouverture, section, obliger, réunir, programme, national...*) et aux principes textuels et scripturaux (*point, virgule, tiret, ligne, flèche, pointillés, parenthèse...*). Les mots représentés par des astérisques, dans le texte de physique suivant, appartiennent tous à la liste du vocabulaire transversal (VT) :

1/* condensateur **capacité C ***** U (figure a) *** o *** armature *** emmagasinée f.** : $C = 10^{-3}$ F : U = 40 V. 2/* condensateur C ***** isolé, *** bobine * inductance L. * résistance * circuit ** négligeable (figure b). ***interrupteur K* instant t = 0. * oscillographe ** visualiser ** u (t) * bornes ** bobine *** courbe *** (figure c). a/** q (t) ***** armature A **date t ***comptée positivement ***** figure. *** courbe *** u (t) ***** maximale **. b/** inductance L ** bobine c/** graphiquement ** instant i (t) *** o* 3,5 ms; d/** emmagasinées ** condensateur *** bobine ** instant t = 0,75 ms.

Ils occupent une place prépondérante (109 mots du VT contre 36 termes ne lui appartenant pas) : sans eux, l'énoncé est incompréhensible. La maîtrise de ces mots (en contexte) peut servir ou desservir les élèves, comme en témoigne l'analyse suivante des productions des élèves.

2 Le français transversal dans les productions d'élèves

Des travaux d'élèves de primaire, réalisés lors de résolution de problèmes mathématiques (RPM), ont été recueillis et analysés du point de vue lexical : ils constituent le deuxième corpus de la présente étude sur le vocabulaire transversal.

Dans le cadre du programme bidisciplinaire mathématiques/français, il a été demandé aux élèves de résoudre des problèmes mathématiques et de fournir, en même temps, des explications détaillées sur le raisonnement, la méthode, la démarche et le résultat obtenu. Chaque élève formule ainsi, par écrit, le processus de sa recherche personnelle, peu importe son résultat. Il contribue, ensuite, à la rédaction d'une fiche commune (tâche par groupes de trois élèves) qui sera présentée oralement, ce qui donnera lieu à des débats et des échanges au sein de la classe tout entière (pour le dispositif pédagogique de RPM, cf. site de l'IREM). De ces verbalisations de la recherche, transparaissent trois ordres de difficultés lexicales.

2.1 Difficultés linguistiques des élèves

L'analyse révèle, outre les problèmes fréquents d'orthographe et de syntaxe qui ne nous intéressent pas ici, trois types de comportements à l'égard du français, en tant que discipline partie prenante de la RPM.

2.1.1 Première difficulté : compréhension de la consigne

Un énoncé mathématique ne comporte pas beaucoup de termes spécialisés, mais les mots courants qui le composent contribuent tous à la construction de son sens. La négligence du moindre mot de la consigne, par exemple, entraîne une incompréhension du problème qui se traduit, forcément, par une réponse inexacte. De même, l'énoncé peut contenir des implicites dont l'interprétation est essentielle pour l'appréhension des informations, de la part des apprenants.

Ainsi, le premier obstacle linguistique (cf. *Annexe 1*), lors de RPM, est créé par le mot, pourtant ordinaire, « chacune », mentionné dans deux phrases de l'énoncé : « [...] 2 pellicules de 20 photos chacune » et « [...] de 12 photos chacune ». Dans les réponses d'élèves, on peut voir, en effet, que l'omission de prendre en compte ce mot fausse complètement la compréhension de l'énoncé et conduit à des raisonnements erronés. Curieusement, ce mot anodin, parfaitement maîtrisé hors du contexte mathématique, échappe à la quasi-totalité des élèves de CM1/CM2 dans deux écoles différentes !

Du côté des auteurs de la consigne, force est de remarquer que la formulation de leur question « Pierre pourra-t-il coller toutes ses photos ? » appelle une réponse par « oui » ou par « non » et n'indique pas explicitement d'effectuer un calcul, ce qui explique peut-être aussi les réponses laconiques et dépourvues de raisonnement des élèves. Ainsi, ceux qui ne savent pas « lire correctement » les énoncés avec leur spécificité ou qui ne savent pas interpréter les questions implicites ont donc du mal à opérer les représentations mentales nécessaires qui permettent de réaliser le calcul ou de mobiliser les procédures exigées par la RPM.

2.1.2 Deuxième difficulté : mise en mots de la RPM

Beaucoup de productions d'élèves se caractérisent par l'absence totale de tout commentaire linguistique et de toute explication concernant la démarche (cf. *Annexe 2*).

De ce point de vue, le respect de la consigne en matière de français n'est pas observé : les élèves n'ont pas joué le jeu de verbaliser leur procédure et de commenter leur résultat de recherche. Dans ce domaine de l'abstraction par excellence, la capacité des élèves, en matière de conceptualisation mathématique, se réalise mentalement sans le recours au langage « naturel ». En conséquence, les élèves ne sont pas à l'aise dans la mise en mots du processus mental ayant présidé à leur RPM. D'autre part, ils n'arrivent pas à concevoir l'intrusion du français dans les mathématiques (et ils le font savoir oralement), d'où le recours exclusif aux représentations sémiotiques : figures, signes mathématiques et symboles divers et variés. La résolution du problème s'effectue donc par formalisation mathématique et sémiotique, sans commentaire linguistique, mais l'illustration joue le rôle d'interprétant intersémiotique. Elle illustre souvent une information représentée par un nom, car les verbes n'ont pas de correspondant iconique. Le langage hybride à l'œuvre, où dessin et opérations mathématiques interagissent, est une stratégie pour pallier l'absence d'expression linguistique. Cette absence de raisonnement linguistique n'est pas spécifique de l'école primaire en FLM. Elle a été observée également dans les classes de lycées en FLS du premier corpus (cf. 1.).

2.1.3 Pertinence des mots utilisés

Certains élèves, qui ont bien compris les enjeux de l'exercice, ne mobilisent que peu de ressources linguistiques parmi celles qu'ils possèdent pour s'exprimer à propos des mathématiques. Leurs productions se caractérisent par une relative indigence lexicale : les connecteurs argumentatifs utilisés sont, par exemple, pertinents, mais peu variés.

Par ailleurs, le genre textuel auquel ils sont habitués (le programme scolaire étant centré sur les textes littéraires et le genre narratif) fait que les élèves privilégient les repères chronologiques au détriment des connecteurs logiques propres aux genres explicatif et argumentatif. Cela explique aussi la répétition du verbe d'action (« j'ai fait », « j'ai fait ») qui met en évidence le processus et les étapes du cheminement intellectuel au détriment du raisonnement qui justifie l'articulation des faits les uns par rapport aux autres.

Toutefois, malgré ces quelques insuffisances linguistiques, leurs textes sont intelligibles ; le professeur peut comprendre le déroulement chronologique des actions, le tâtonnement et les mouvements de leur pensée successifs. L'enchaînement des énoncés explique la logique de leur raisonnement et justifie le résultat obtenu ainsi que la démarche adoptée. L'avantage de cet exercice est justement de connaître où se situe exactement l'obstacle, si obstacle il y a, pour y remédier ultérieurement, dans une éventuelle intervention pédagogique. Les mots suivants, appartenant tous au VT de notre liste, sont attestés dans les textes d'élèves (cf. Annexe 3) : – Verbes : faire, trouver, aller, mettre, chercher, essayer, rester, former, contenir, se rendre compte, comprendre, dire, compter, enlever, rajouter, avoir, prendre – Noms : calcul, résultat, essai, cartes, côté, carré, triangle – Adjectifs : supérieur, plus petit

que – Connecteurs chronologiques (*au début, puis, et, après, ensuite, encore*); et connecteurs logiques (*parce que, sachant que, et, ça fait, mais, donc*), etc.

On voit bien qu'un élève qui justifie son opinion, qui argumente ou qui formule des résultats de recherche en mathématiques, fait preuve d'une capacité autant de construire une connaissance propre en mathématiques, que de construire une compétence transversale puisqu'ici le savoir passe aussi par le langage.

2.2 Compétences requises en FT pour la résolution de problèmes

Le français est un vecteur de compréhension et d'expression en mathématiques. C'est pourquoi la RPM qui occupe une place de choix dans le programme de cette discipline est révélatrice des compétences et des lacunes des élèves en matière de français. Les textes institutionnels (MEN 2002 : 15) pointent la difficulté des élèves à lire les énoncés et à comprendre certains mots polysémiques. Afin d'améliorer leur compétence langagière, l'on doit limiter les erreurs liées au français : elles sont, en partie, responsables des mauvaises notes en mathématiques et des échecs scolaires en général.

1. Les représentations sémantiques erronées sont parfois induites par la polysémie des termes comme : *point, milieu, multiple, sommet*. Ainsi, le « sommet » d'un triangle en géométrie n'est pas nécessairement « en haut », or « sommet » (sens général) évoque « haut » ; le mot quelconque renvoie à une forme spécifique du triangle (et non à « n'importe quel » triangle).
2. Certains élèves sont incapables d'opérer les inférences indispensables à l'interprétation des données. Les énoncés de problèmes arithmétiques sont nécessairement lacunaires puisque le choix de l'opération est le véritable enjeu de la résolution. Il est lié à l'identification des relations entre les données ; or, ces relations ne sont pas évidentes pour les élèves. On l'a vu avec les réponses par « oui » ou par « non » (*cf. Annexe 2*).
3. Les références de certains mots (*fois, partage, reste, différence, total...*) induisent la mobilisation de notions ou de procédures qui ne sont pas toujours bien maîtrisées, par exemple certains verbes suscitent deux types de tâches difficiles à distinguer : matérielles (*relever, encadrer*) ou intellectuelles (sens abstrait de « mettons que », ou de *ranger, développer, compléter*).
4. Certains élèves ont la propension à ne s'occuper que des nombres fournis par l'énoncé, sans se soucier des autres mots. Le choix de ces éléments mathématiques au détriment de leurs entourages linguistiques reflète la désinvolture des élèves ou leur reproduction inconsciente de règles imposées antérieurement ; par exemple, le cours précédent porte sur la division, alors le problème sera résolu par la division : signe que la finalité de l'exercice n'est pas comprise. Cette représentation de la langue est illustrée par le titre du livre de Stella Baruk, *L'âge du capitaine*, en référence au problème donné aux élèves (« dans un bateau, il y a 26 moutons et 10 chèvres ; quel est l'âge du capitaine ? »), et aux réponses de type : « $26 + 10 = 36$ ans ! »

En résumé, les difficultés à résoudre les problèmes sont de deux ordres : la compréhension de la situation évoquée par les mots de l'énoncé et la capacité à mobiliser les acquis. Le programme souligne qu'il est particulièrement important que les élèves sachent naviguer entre données et questions, passer du texte à d'autres formes de représentations, interroger leurs acquis, mobiliser leur connaissance, etc. Ces compétences de navigation intersémiotique relèvent-elles de la transdisciplinarité ?

3 Pédagogie et didactique de la transdisciplinarité

3.1 Mise en place progressive de l'inter-/transdisciplinarité dans le système éducatif

Le « français dans les autres disciplines » ne devient, que depuis peu, objet de préoccupation dans l'enseignement en France, bien que toutes les Instructions officielles successives, depuis 1970, évoquent la question de la langue au sein des disciplines scolaires.

On a d'abord imputé l'échec scolaire, dans les disciplines scientifiques, aux jargons jugés prétentieux et aux protocoles considérés comme ésotériques : on a pensé que le langage spécialisé était susceptible d'être enseigné à part. Ensuite, les représentations ont évolué, ainsi que les prescriptions institutionnelles : on considère alors comme un paradoxe de prétendre former des citoyens par des actions parallèles conduites séparément. En liant les matières enseignées, on entend pallier la parcellisation des connaissances juxtaposées. Mais la question de l'interdisciplinarité rencontre des résistances, car chaque spécialité tient à sa souveraineté.

Entre les années 1990 et 2000, le concept de « transdisciplinarité » (trans signifie : « entre/à travers/au-delà des disciplines », d'après la Charte de la transdisciplinarité, 1994) émerge à tous les niveaux d'enseignement et prolifère dans les programmes sous des étiquettes différentes : « décloisonnement », « maîtrise de la langue », « français dans les autres disciplines », « compétence/attitude/connaissance/savoir fondamental(e) » (cf. Armand 1999).

Le programme 2008, recentré sur « les fondamentaux », se veut « le 1^{er} étage de la fusée « réussite pour tous » ». C'est pour cela, sans doute, que le Socle commun (MEN 2006) accorde une place de choix à « la maîtrise de la langue » dont la dialectique inévitable avec le savoir implique que la langue tire bénéfice des apprentissages disciplinaires et vice-versa. Les programmes mettent en place, par ailleurs, de nouveaux objets didactiques : au primaire (mathématiques/français : cf. 2. RPM) ; au collège (l'éducation civique, juridique et sociale (ECJS), l'éducation à l'environnement et au développement durable (EEDD) et l'itinéraire de découverte (IDD)) ; et au lycée (travaux personnels encadrés (TPE), sans oublier la science en langue étrangère pour le cas des sections européennes).

Qu'il s'agisse du « trans- » ou de « l'inter-disciplinaire », le « décloisonnement » se pose de façon cruciale dans le système éducatif.

3.2 Décloisonnement inter- et intradisciplinaire : quel savoir-faire en français ?

Considérée tantôt comme une condition, tantôt comme un prolongement des apprentissages, la transversalité prônée par les nouvelles directives ne bénéficie, cependant, que rarement d'un accompagnement didactique, d'où la liberté (ou l'embaras) des professeurs dans la conception de leurs projets polydisciplinaires.

Les disciplines enseignées à l'école s'opposent habituellement les unes aux autres, bien qu'elles soient inséparables de la langue et que leur compréhension soit liée à l'appréhension de l'outil linguistique qui les constitue. Cette interaction parties/tout relève de la « pensée systémique » qu'E. Morin (1999) qualifie de « défi » (défi de la globalité). Elle permet de relier ce qui est pourtant « tissé ensemble », souligne-t-il. Les progrès nouveaux tendent donc à transgresser les frontières historiques des disciplines : les interactions et la complexité remplacent la disjonction des savoirs qui « fractionne les problèmes ».

Dans les programmes, le sens du « français dans les autres disciplines » ne va pas de soi. Car ce facteur qui participe à la construction des savoirs n'est pas clairement défini. Le cloisonnement traditionnel du français ne réside pas seulement dans la distinction classique entre langue et littérature, mais dans le contexte d'acquisition fort complexe reflétant deux images de la langue (générale et scientifique) entre lesquelles, se situe, parfois, le FLE. À cela, s'ajoute la dichotomie : terminologie/langue transversale qui émerge à l'intérieur même de la langue maternelle, la confrontant à elle-même dans ses spécifications contextuelles. Les différentes catégories peuvent se chevaucher en créant des difficultés de positionnement. Avant donc de faire se rencontrer les disciplines, nécessité est de travailler l'ouverture de chacune d'elles, à commencer par le français lui-même.

Conclusion

Sur le plan pédagogique, guider les apprentissages en adoucissant le babélisme et les frontières disciplinaires revient, selon nous, à faire acquérir, d'abord, le vocabulaire commun qui traverse la quasi-totalité des disciplines, et à sensibiliser à son importance pour la réussite scolaire. Chaque enseignant, quelle que soit sa spécialité, est susceptible de faire découvrir la langue derrière les discours techniques et faire évoluer les représentations et les enjeux des savoirs disciplinaires. Les problèmes concrets de vocabulaire peuvent être abordés en toute circonstance pour amener l'élève à se familiariser avec les composantes linguistiques du français scientifique : terminologie, néologie, vocabulaire général, français élémentaire, etc., et à distinguer les divers genres de textes scientifiques : discours d'experts, discours de vulgarisation, discours pédagogique scolaire : le manque de repères est une source d'échec.

Sur le plan linguistique, l'étude présentée ici expose les problèmes théoriques et méthodologiques de l'extraction du vocabulaire transversal, un champ du savoir à l'intersection de la terminologie et de la langue générale susceptible d'intéresser les lexicologues, terminologues et traducteurs, mais aussi les scientifiques et les

didacticiens, car il n'est pas seulement « passeur de mots », mais aussi passeur du savoir, donc « passeur d'espoir ».

Bibliographie

- Armand (A), 1999 : « Curieuse situation de la discipline Français », dans *Les Cahiers de L'EPS de l'Académie de Nantes* n° 20, juin, p. 35-37.
- Charte de la transdisciplinarité, 1994 : *Premier congrès mondial de la transdisciplinarité*, 26 novembre, Portugal, « Convento da Arrabida », dernière mise à jour le 24 janvier 2002, basarab.nicolescu.perso.sfr.fr/ciret/chartfr.htm.
- Chetouani (L.), 1997 : *Vocabulaire général d'enseignement scientifique*, Paris, L'Harmattan.
- Coxhead (A), 2000 : « A new academic word list », dans *TESOL Quarterly*, 34-2, p. 213- 238.
- Drouin (P), 2007 : « Identification automatique du lexique scientifique transdisciplinaire » dans *Revue française de linguistique appliquée*, n° XII-2, p. 45-64.
- Gougenheim (G.), et al. : 1959, 1964 : *Élaboration du français élémentaire*, Paris, Didier.
- Juilland (A.) et al., 1971 : *Frequency Dictionary of French Words*, Paris, Mouton.
- Kocourek (R.), 1982 : *La langue française de la technique et de la science*, Paris, La Documentation française.
- Lerat (P), 1995 : *Les langues spécialisées*, Paris, PUF.
- Martin (R.) et al., 1971 : *Dictionnaire des fréquences du Trésor de la langue française*, Paris, Klincksieck
- Ministère de l'Éducation nationale, 2006 : *Le socle commun des connaissances et des compétences*, Paris, CNDP.
- Ministère de l'Éducation nationale, 2002 : *Document d'accompagnement du programme. Cycle 3*, Paris, CNDP.
- Morin (E.), 1999 : *Relier les connaissances. Le défi du XXI^e siècle*, Paris, Seuil.
- Phal (A.), 1971 : *Vocabulaire général d'orientation scientifique*, CREDIF/Didier.
- Tutin (A), 2007a : « Traitement sémantique par analyse distributionnelle des noms transdisciplinaires des écrits scientifiques », dans *Actes de TALN 2007*, Toulouse, 5-8 juin, p. 283-292.
- Tutin (A), 2007b : « Autour du lexique et de la phraséologie des écrits scientifiques » dans *Revue française de linguistique appliquée*, n° XII-2 (présentation).

Annexe 1

Les photos
 Au cours d'un voyage, Pierre a d'abord utilisé deux pellicules de 20 photos chacune. Il a constaté que 4 photos n'étaient pas réussies.
 Ensuite, il a utilisé deux nouvelles pellicules de 12 photos chacune et cette fois, deux photos ont été ratées.
 Pierre colle les photos réussies dans un album de 10 pages; sur chaque page on ne peut pas coller plus de 6 photos.
 Pierre pourra-t-il coller toutes les photos réussies dans son album ?

Ma recherche

$20 - 4 = 16$
 $12 - 2 = 10$
 $16 + 10 = 26$
 $26 \div 6 = 4 \text{ R } 2$

Ma réponse

26

Ma recherche

$20 \div 4 = 5$
 $12 \div 2 = 6$
 $5 + 6 = 11$
 $11 \times 6 = 66$

$1 + 6 + 6 + 6 + 6 + 6 + 6 + 6 = 53$
 $30 + 30 = 60$
 $20 + 20 + 20 = 60$
 $10 + 10 + 10 + 10 + 10 = 50$

Ma recherche

$26 - 6 = 20$
 $20 - 6 = 14$
 $14 - 6 = 8$
 $8 - 6 = 2$
 $2 - 6 = 0$

Ma réponse

Oui il peut placer toutes les photos.

Annexe 2

Résous le problème 1
 Le mobilier de l'école
 Une entreprise a acheté trois chargements de 300kg chacun pour équiper en mobilier une école.
 Le premier chargement contient 15 tables et 30 chaises.
 Le second contient 25 tables.
 Le troisième contient 10 tables, 20 chaises et 5 armées.
 Combien pèse une chaise, une table, une armée ?

Ma recherche :

$9 \times 25 = 225$
 $12 \times 30 = 360$
 $180 + 120 = 300$
 $120 \div 4 = 30$
 $30 \times 4 = 120$
 $100 \div 4 = 25$
 $20 \div 4 = 5$
 $100 \div 10 = 10$
 $100 \div 20 = 5$
 $100 \div 5 = 20$

PROBLEME

Un fermier possède des poulets et des lapins dans sa basse-cour.
 On compte 63 têtes et 230 pattes. Combien de poulets et de lapins possède-t-il ?

Première recherche

Ma recherche et mes explications :

$63 \times 2 = 126$
 $230 - 126 = 104$
 $104 \div 2 = 52$
 $52 \times 4 = 208$
 $208 + 52 = 260$
 $260 - 230 = 30$
 $30 \div 2 = 15$
 $15 \times 4 = 60$
 $60 + 15 = 75$

Réponse

Annexe 3

Prestations de groupes d'élèves :

(Enoncé sur le mobilier de l'école) :

Exemple 1 :

On a commencé par le deuxième chargement où il n'y avait que des tables. On a calculé douze fois vingt-cinq et on a trouvé 12 kg. On a continué par le premier chargement. Il y avait des chaises et des tables. On a enlevé le poids des tables et on a trouvé 4 kg. On a fini par le troisième chargement. Il y avait des tables, des chaises et des armoires. On a enlevé le poids des tables et des chaises.

Exemple 2 :

Nous avons commencé par le second chargement pour trouver combien pèse une table. On a fait 300 divisé par 25 et on a obtenu 12 kg. Ensuite, nous avons fait le premier chargement parce qu'il y avait des tables et des chaises, alors, on a enlevé le kilo* des chaises au camion et on a fait $15 \times 12 = 180$. On a enlevé à 300 180 et on a obtenu 120. Alors, on a fait 120 divisé par 30 et on a obtenu 4 kg ; ensuite on a fait le troisième chargement parce qu'il avait 10 tables, 20 chaises et 5 armoires et comme on savait le kilo* des tables, on a fait $10 \times 12 = 120$, et comme on savait le kilo* des chaises, on a fait $20 \times 4 = 80$. Après, on a additionné dans notre tête $120 + 80 = 200$. On a enlevé 200 à 300 et ça fait 100 divisé par 5 et on a obtenu 20 kg.

Alors les tables pèsent 12 kg, les chaises pèsent 4 kg et les armoires pèsent 20 kg.

NB :* on peut noter ici la confusion entre : « kilo » et « poids », et la distinction entre « kilo »/ « kg », ce dernier correspondant bien à l'unité de mesure.

Prestations individuelles :

Enoncé du problème « pour chercher » : On a 18 formes qui sont des carrés et des triangles. Au total, le nombre de côtés de toutes ces formes est 60. Combien de carrés ? Combien de triangles ?

Elève 1 :

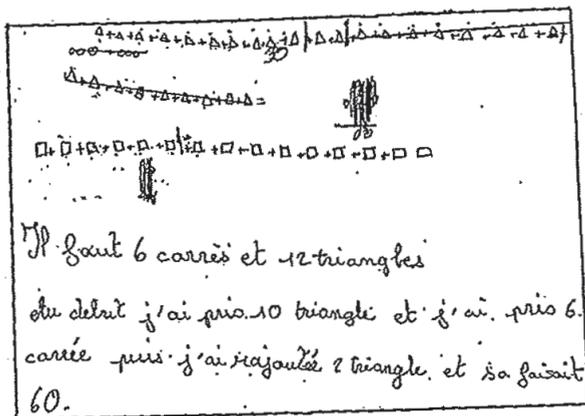
J'ai fait plein de calculs qui ne m'ont pas servi beaucoup. Après, j'ai cherché combien et combien était = à 60 et j'ai trouvé $36 + 24$ parce que c'était = à 60. J'ai fait $9 \times 4 = 36$ et $8 \times 3 = 24$. Ça faisait 60 mais pas 18 certes et puis après j'ai fait $3 \times 12 = 36$ et $6 \times 4 = 24$; ça fait 60 et ça faisait 18 cartes. J'avais trouvé.

Elève 2 :

J'ai fait 12 triangles (en sachant qu'un triangle contient 3 côtés) : $12 \times 3 = 36$. Il y a déjà 12 formes et 36 côtés. Il reste 6 formes et 2 côtés à mettre. Un carré contient 4 côtés : $6 \times 4 = 24$.

Elève 3 :

Il faut 6 carrés et 12 triangles. Au début, j'ai pris 10 triangles et j'ai pris 6 carrés, puis j'ai rajouté 2 triangles et ça faisait 60.



Christine Jacquet-Pfau,
Collège de France et LDI,
Paris,
France,
ch.jacquet-pfau@college-de-france.fr.

Jean-François Sablayrolles,
Paris 13 et LDI,
Paris,
France,
jfsablayrolles@wanadoo.fr.

John Humbley,
Paris 7 et LDI,
Paris,
France,
humbley@eila.univ-paris-diderot.fr.

Emprunts, créations « sous influence » et équivalents

Résumé

Outre un champ indiquant la matrice lexicogénique (interne ou externe pour les seuls vrais emprunts, directs ou adaptés) du néologisme d'une fiche de la base néologique Neologia, deux autres champs notent une éventuelle influence étrangère reconnaissable dans la création de néologismes « français » (matrice interne) et la langue exerçant cette influence. Faux emprunts, structure régressive dans des composés de type germanique, hybrides, traduction, calques et créations équivalentes plus libres relèvent de matrices internes. Ce sont des emprunts relevés récemment et leur(s) équivalent(s) français, néologismes sous influence, que nous décrivons ici et dont nous comparons la vitalité respective.

Mots-clés : emprunt ; néologisme ; équivalent ; calque ; mesures de diffusion/circulation.

Introduction

Dans le cadre d'un projet de coopération avec l'Université Aristote de Thessalonique portant sur la comparaison des emprunts et de leur traitement en grec moderne et en français, l'équipe néologie du laboratoire LDI (« Lexiques, Dictionnaires et

Informatique » des universités Paris 13 et Cergy-Pontoise) a étudié les emprunts entrés dans sa base de données, *Neologia*, ainsi que les emprunts communiqués par l'équipe grecque et circulant en français. Les données du corpus ont été extraites de la presse entre 2006 et juin 2008 essentiellement, mais leur première attestation est souvent bien antérieure. Nous précisons les fondements théoriques, les objectifs et la méthode de travail de cette étude. Puis nous en présenterons les résultats français et esquisserons quelques perspectives.

1 Assises théoriques, objectifs et méthode du projet

1.1. Assises théoriques

Le terme *emprunt* pose de nombreux problèmes et il est employé pour des faits multiples et disparates. Pour un éclaircissement sur ce sujet voir, entre autres, l'article de Esme Winter-Froemel (2009) et son abondante bibliographie.

Parallèlement à cette réflexion, mais indépendamment car antérieurement, nous avons distingué, pour le traitement des néologismes dans la base *Neologia*, les néologismes par emprunt (direct) à une autre langue et plusieurs types de créations de mots par des matrices internes du français mais sous l'influence d'une langue étrangère. Ce faisant, nous nous inscrivons dans la continuité des travaux de Jean Tournier (1985 et 1991) et le tableau des matrices mis en œuvre au LDI (v. tableau 1) est fortement tributaire de ceux qu'il avait proposés où la matrice externe qu'est l'emprunt s'oppose globalement à l'ensemble des matrices internes.

En sus de la matrice externe qu'est l'emprunt et qui figure au sein du tableau des matrices lexicogéniques (tableau 1), nous avons donc inclus, dès la création de la base de néologismes *Neologia*, deux champs réservés au mode d'influence (traduction, calque, utilisation d'un formant d'origine étrangère, utilisation d'une structure d'origine étrangère, trou à combler¹) et à la langue qui l'exerce, l'anglais le plus fréquemment, pour des néologismes créés par des matrices internes. Il s'agit de pouvoir traiter comme il convient les faits suivants, assez bien connus, mais pas nécessairement analysés de manière satisfaisante, puisque ce ne sont pas des emprunts à proprement parler. On peut en distinguer plusieurs grandes catégories.

1 Création pour combler une lacune révélée dans le lexique d'une langue par comparaison avec une autre langue.

| | | | | | | |
|-------------------|-----------------------------------|--|-----------------------|---|---|--------------------------------------|
| matrices internes | morphosémantiques | construction | affixation | préfixation | <i>détatouer</i> | mscapr |
| | | | | suffixation | <i>Statuesque</i> | mscasu |
| | | | | dérivation inverse | <i>prester</i> | mscadi |
| | | | | parasynthétique (?) | <i>désidéologisé (?)</i> | mscapa |
| | | | flexion | | <i>ils closirent,</i> <i>la représaille</i> | mscfech |
| | | | composition | composition synapsie composition savante hybride | <i>voiture-bélier</i> <i>lanceur d'alerte</i> <i>batracianophile</i> <i>e-commerce,</i> <i>aquacinéaste</i> | mscco msccsy msccsav mscchy |
| | | | | mot valise | <i>peopolitique</i> | msccmv |
| | | | | compoction | <i>mobinaute, dircab</i> | msccct |
| | | | | onomatopée fausse coupe paronymie | <i>dzoing</i> <i>la nethésie</i> <i>infractus</i> | msidon msiddef |
| | | | syntacticosémantiques | changement de fonction | conversion | <i>la glisse, la gagne</i> |
| | conversion verticale | <i>de rejuvenation</i> | | | ssfconv | |
| | déflexivation | <i>le boire, le manger</i> | | | ssfdef | |
| | combinatoire syntax./ lexicale | <i>ça craint</i> <i>la prise de train</i> | | | ssfcosy ssfcolex | |
| | changement de sens | métaphore | | <i>souris (inform.)</i> | sssmph | |
| | | métonymie | | <i>transistor (poste)</i> | sssmny | |
| | | autres figures | | <i>partir « mourir »</i> | sssfgr | |
| | | troncation | | <i>blème, petit déj</i> | mretr | |
| | morpho- logiques | réduction de forme | siglaison/acronyme | <i>LMD, ECLUE</i> | mresi | |
| | | | détournement | <i>planche à promesses</i> | prdet | |
| | pragmatico-sémantique | | | | | |
| matrice externe | | | emprunt | <i>binge drinking</i> | exemp | |

Tableau 1 : tableau des matrices lexicogéniques

Le premier ensemble contient les faux emprunts, c'est-à-dire des mots qui n'existent pas dans la langue censée être la langue source. Cet ensemble regroupe :

1. des mots formés avec des éléments d'une même langue étrangère (*best-thriller*, *boomer bohemian*, *dograpper*², *food-trotter*...);
2. des hybrides, adoptant souvent la structure régressive germanique (*Atys-boomer*, *bommer-bohème*, *gay attitude*, *girly attitude*³, *pipolitique*, *serial menteur*);
3. des mots dont l'origine des formants est moins facilement déterminable, puisque la forme serait identique dans plusieurs langues (*détriangler*⁴, *hamburgerisation*, *overdosé* (n.m.), *victimation*...);
4. des influences de structure (les composés d'ordre régressif avec des éléments autochtones du type *royale attitude* ou mixtes comme en ii).

2 Sans doute récréation indépendante de son existence en anglais. Un internaute se flatte même d'avoir créé le mot.

3 Ces deux derniers éléments sont beaucoup plus fréquents en français qu'en anglais, qui pourrait bien avoir importé ces faux emprunts.

4 Au sens politique, contraire de *triangler* « s'emparer des valeurs de son adversaire » emprunté à Tony Blair qui l'a fait connaître.

Le second ensemble est formé des cas de traduction (*souris/mouse* dans le domaine informatique).

Le troisième, assez proche du cas précédent mais qui en diffère dans la mesure où il met en jeu des unités polylexicales, comprend les calques (*on line/en ligne, product placement/placement de produit*).

Le quatrième ensemble regroupe des créations équivalentes plus libres (*binge drinking/cuite express* ou *biture express*⁵; *flash over/embrasement généralisé éclair*) qui servent à combler une lacune dans le lexique de la langue cible.

Le cinquième ensemble comprend des néologismes sémantiques sur des mots qui ont été empruntés (*flasher* « prendre en flagrant délit d'excès de vitesse »).

Un sixième ensemble est constitué des mots créés sur la base d'un emprunt (*boboïser*...).

Toutes ces créations relèvent de matrices lexicogéniques internes, la composition (*embrasement généralisé éclair*), avec ses variantes que sont les mots-valises (*pipolitique*) et les synapsies (*placement de produit*); la suffixation (*boboïser*); la conversion (verticale) d'une séquence syntaxique (un syntagme prépositionnel) en unité lexicale (*en ligne*); la néologie sémantique (*souris*); etc. Bien des matrices différentes peuvent être utilisées dans ces créations internes « sous influence ».

1.2 Objectifs

L'objectif premier est la comparaison des faits d'emprunt en grec et en français, mais les considérations précédentes conduisent à ne pas s'en tenir au seul établissement de listes d'emprunts directs à mettre en regard. S'en tenir à celles-ci pourrait même conduire à des conclusions erronées. Pour être efficace, la comparaison doit se donner des objectifs plus précis, qui se ramènent à deux exigences principales.

La première de ces exigences réside dans la prise en compte des créations « sous influence », en particulier des traductions, calques et équivalents. Les faux emprunts, les influences de structure, et les utilisations de formants étrangers sont moins pertinents pour cette recherche : aussi ne les avons-nous pas retenus dans notre corpus. Seuls les emprunts directs et leurs équivalents/solutions de remplacement ont de l'intérêt pour ce travail comparatif interlangue.

La seconde exigence réside dans la prise en compte des degrés de circulation respectifs d'un emprunt direct dans une langue et de la ou des formes équivalentes dans la même langue. *Docudrama* et *on line* figurent dans la liste des emprunts du grec à l'anglais communiqués par l'équipe grecque. On relève en français des occurrences de ces deux emprunts. Mais s'en tenir à cette seule constatation que le français et le grec ont fait les mêmes emprunts oblitère complètement la réalité linguistique française.

5 Remarquons que, si ces mots désignent la même réalité, l'équivalence n'est pas parfaite puisqu'à la notion de quantité massive de l'anglais correspond en français la notion de rapidité de l'absorption d'alcool.

L'emploi de ces emprunts est fortement minoritaire face à leurs concurrents français. La création française *docufiction* est beaucoup plus représentée que l'emprunt *docudrama* (28 occurrences contre 3 dans des archives de *Libération*⁶, et 119 000 pages de *Google* contre 2 880, en juin 2008). De même il y a 856 occurrences du français *en ligne* dans *Libération* contre 39 de l'emprunt anglais *on line*, disproportion confirmée, mais dans une moindre proportion, par le nombre de pages dans *Google* : 4 160 000 contre 680 000, au moment de l'interrogation, toujours en juin 2008.

1.3 Méthodologie

Les données précédentes donnent des indications sur la méthode utilisée. Nous allons les expliciter puis donner quelques précisions supplémentaires sur la composition du corpus.

1.3.1 Des mesures comparatives de circulation

Pour comparer la vitalité des emprunts et de leurs équivalents, nous avons d'abord conduit un travail lexicographique en relevant dans des dictionnaires monovolumaires récents (éditions 2009 des *Petit Larousse*, *Petit Robert* et *Hachette*) et dans l'encyclopédie en ligne *Wikipedia* (certes moins « autorisée » mais beaucoup plus réactive) la présence ou l'absence des uns et des autres ainsi que les marques d'usage et renvois à d'autres articles ou recommandations officielles.

Nous avons également mesuré leur circulation à partir de requêtes sur des archives de quotidiens généraux nationaux (*Libération* essentiellement) et en utilisant le moteur de recherches *Google* dont les résultats peuvent être instructifs, moyennant quelques précautions d'interprétation des résultats chiffrés. Il faut en effet prendre garde à l'existence d'homonymes, en particulier de noms propres, à la multiplication des renvois à une même occurrence originelle, etc. Des disproportions numériques importantes, quand on compare des chiffres élevés, peuvent néanmoins être l'indice d'une différence de diffusion. Des résultats peu nombreux sont assurément la marque d'une absence de grande diffusion.

Ces données synthétisées ont été intégrées dans un tableau à neuf colonnes, la première pour le mot source, puis quatre colonnes pour le grec et autant pour le français : la forme de l'emprunt dans la langue cible, s'il existe⁷, puis les informations lexicographiques, la circulation dans *Libération* et *Google*, enfin l'équivalent ou les équivalents dans la langue cible. Un deuxième fichier texte, plus volumineux, contient l'essentiel de ces informations ainsi que les définitions et des contextes d'emploi.

1.3.2 Remarques sur la constitution du corpus

Pour cet article, nous restreignons notre corpus à des emprunts récents (une quinzaine d'années). Ne sont pas pris en compte les emprunts qui se sont installés dans

6 Consultées par John Humbley en juin 2008.

7 Et s'il n'existe pas, une forme est inventée, suivie de l'astérisque signalant la non existence, ce subterfuge ayant pour objectif de maintenir l'ordre alphabétique des données.

la langue depuis longtemps, même s'ils gardent des spécificités graphiques ou morphologiques comme *week-end*. Ont ainsi été exclus des anglicismes comme *tweeter* ou *woofer*, transmis par l'équipe grecque, car introduits en français depuis les années 1950. Il en va de même de *peep show*, daté de 1980, que la recommandation officielle de diffusion très limitée *mirodrome* n'a pas remplacé, ou encore de *trash*, daté de 1989 dans *PR*.

Ont été éliminés également les emprunts donnés par l'équipe grecque dont nous n'avons pas trouvé trace en français, comme *after dinner*, *stag*, etc., ou dont nous ne savons pas avec quel sens ils existent en grec : *fusion* (musique ? cuisine ?).

N'ont été retenus que des emprunts et leurs éventuels équivalents qui commencent à circuler : sont donc exclus les hapax (*Atys-boomer*, *food-trotter...*, qui ne sont mentionnés ici que comme parfaites illustrations de créations françaises à l'aide de formants étrangers) et les xénismes (*agogo*, *bata*, *shekere*, instruments de musique d'origine africaine, figurant dans la liste des emprunts grecs).

Enfin, nous avons privilégié pour le présent travail des emprunts et néologismes qui sont révélateurs de faits de société émergents. Les éléments de ce corpus, au sein duquel figureront des emprunts sans équivalent (*smirting*), ou avec un seul équivalent (*blook/blouquin* ; *junk food/malbouffe* ; *watermarking/tatouage numérique*) ou plusieurs équivalents (*phishing/hameçonnage/filoutage* ; *bling-bling/nouveau riche/tape-à-l'œil*) doivent, du fait de l'internationalisation grandissante d'un grand nombre de phénomènes, se retrouver dans beaucoup d'autres langues, soit sous forme d'emprunts, soit sous forme de créations propres, soit sous les deux formes, avec la prédominance de l'une des deux.

2 Emprunts et équivalents français : les résultats

La présentation des données se fait d'une manière ordonnée, en distinguant quatre ensembles d'inégale dimension selon qu'un emprunt n'a pas d'équivalent, a un seul équivalent, a plusieurs équivalents ou, cas rare, qu'un équivalent existe sans que l'emprunt soit attesté. Ce sont en tout cent soixante-trois formes, emprunts ou équivalents, qui constituent le corpus d'étude. Puis nous examinerons la typologie des équivalents.

2.1 Circulation relative des deux

Dans deux cas, il n'y a pas de circulation relative à mesurer puisqu'une seule des deux formes existe. C'est nettement plus souvent celle de l'emprunt direct, mais une fois, au moins, il y a un équivalent sans que la source soit attestée. Quand un ou des équivalents existent, il n'est pas toujours facile de mesurer la diffusion relative, pour diverses raisons (homonymie, nombres d'attestations trop faibles pour être significatifs...), mais on a parfois des disproportions significatives. Il faut aussi tenir compte des recommandations officielles, au succès variable.

2.1.1 Emprunts sans équivalents

Les emprunts sans équivalent sont au nombre de quarante-sept. Quarante sont assurés, dont trois avec des aménagements graphiques. Il y a également cinq cas indécidables et deux cas complexes.

Liste des quarante cas assurés (dont trois avec aménagement graphique):

| | | |
|----------------------------------|--|-------------------------|
| <i>bobo</i> ⁸ | <i>gender generation gap</i> | <i>smirting</i> |
| <i>chick movie</i> | <i>in game</i> | <i>smoky eye</i> |
| <i>cine-business</i> | <i>legging</i> | <i>spin off</i> |
| <i>comarketing</i> | <i>low rider</i> | <i>spoof</i> |
| <i>cyberlaundering</i> | <i>low riding</i> | <i>spoofers (verbe)</i> |
| <i>daily strip</i> | <i>MVNO</i> | <i>spoofing</i> |
| <i>donk style</i> | <i>news factory</i> | <i>stollen</i> |
| <i>endorsement</i> | <i>people/pipole</i> ⁹ | <i>triangulation</i> |
| <i>facebooker</i> | <i>pharming party</i> | <i>triangler</i> |
| <i>fan subbing</i> | <i>quirk</i> | <i>typosquatting</i> |
| <i>free hug</i> | <i>rogue trader</i> | <i>undercover</i> |
| <i>gatecrasher</i> ¹⁰ | <i>silicone free (dom. chirurgie esthétique)</i> | <i>USB</i> |
| | | <i>web radio</i> |

et avec aménagement graphique:

chèque book journalisme
drog stop
fixeur

Liste des cinq cas indécidables:

dognapper
freenaute
spameur
spoofeur (avec aussi la graphie anglaise *spoofers*)
trash attitude.

Il est impossible de décider s'il s'agit d'emprunts ou de créations françaises indépendantes de l'existence du mot en anglais. C'est particulièrement clair dans le cas de *dognapper*, dont les anglophones disent qu'il est attesté depuis les années

8 Emprunt (2000 d'après PR) sur la base duquel a été développée toute une série de dérivés par suffixation: *boboiser*, *boboisation*, *boboïsant*, *boboïste*...

9 Sur l'emprunt *people* (1988 d'après PR), avec les graphies *pipole* et *pipeule* a été développée toute une série de dérivés par suffixation: *pipolerie*, *pipoliser*, *pipolitisation*, *pipolitique*, avec une ou plusieurs graphies pour la première syllabe *pip-/peop-*.

10 Les équivalents parfois proposés *s'incruster*, *resquiller*, *être un pique-assiette* ne correspondent pas exactement au fait de s'imposer dans des soirées festives privées sans invitation.

quarante (dans des films de Walt Disney), mais qu'un internaute français se vante sur son blog d'avoir créé. L'occurrence trouvée dans le quotidien gratuit 20" est probablement une autre recreation d'un journaliste.

Pour *freenaute* et *trash attitude*, il se trouve que le formant (d'origine grecque) *-naute* ou la structure *X attitude* sont bien plus utilisés en français qu'en anglais, où les quelques occurrences pourraient bien être des emprunts... au français (de faux anglicismes!).

La constitution d'une famille de mots, en particulier de noms d'agent à partir de noms d'action, pour *spammeur* et *spoofeur* par exemple, peut très bien s'effectuer directement en français, ou être le fruit d'emprunts, francisés ou non. Et, selon les occurrences, les deux cas doivent se présenter sans que ce soit décidable.

Il y a également deux cas complexes : deux emprunts probablement au chinois de faux anglicismes, pour une pratique apparue en Chine avec plusieurs attestations dans des journaux français à partir de mars/avril 2008 :

team buyer
team buying.

Les calques *achat en groupe* et *acheteur en groupe* ne correspondent pas à cette pratique innovante où des individus qui ne se connaissent pas se donnent rendez-vous en nombre, via internet, devant un magasin pour faire baisser les prix d'un produit qu'ils désirent tous acheter. Les termes français existent, mais ils dénomment des regroupements de professionnels détaillants, afin d'obtenir des prix d'achat identiques à ceux de la grande distribution.

2.1.2 Emprunts avec un équivalent

La somme des emprunts et de leur(s) équivalent(s) atteint soixante-quinze, un chiffre impair, car un équivalent est utilisé pour deux emprunts différents, correspondant à des réalités différentes. Il y a donc trente-sept équivalents pour trente-huit emprunts. Les données sont présentées sous la forme d'un tableau à cinq colonnes (tableau 2). La première est réservée à l'emprunt et la seconde à son équivalent. Les trois autres colonnes sont consacrées à des analyses de l'équivalent. Dans la troisième colonne figure une analyse morphologique en préfixe, radical, suffixe¹¹, afin de bien faire la distinction avec la matrice lexicogénique interne qui a créé le néologisme et qui est indiquée dans la quatrième colonne. La cinquième colonne est consacrée au mode d'influence exercé par la langue source sur la langue cible (le français en l'occurrence).

11 Nous avons également utilisé les notations suivantes : *cc* pour conjonction de coordination, *prep* pour préposition et *let* pour les constituants de sigle. *Conversion v* dénomme une conversion verticale. Un trait d'union est utilisé pour les formants d'une unité monolexicale et un blanc sépare les constituants d'unités polylexicales.

| Emprunts | Équivalent | Configuration | Matrice | Mode |
|--------------------------------|------------------------------------|--------------------------|----------------|-------------|
| <i>blog/blogue</i> | Recom. off. : bloc-notes | rad rad | sémantique | trou comblé |
| <i>blogueur</i> | carnetier (rare) | rad-suff | suffixation | trou comblé |
| <i>blook</i> | blouquin (rare) | rad-rad | mot-valise | calque |
| <i>cobranding</i> | co-marquage | préf-rad-suff | préfixation | trou comblé |
| <i>cosponsoring</i> | parrainage | rad-suff | sémantique | trou comblé |
| <i>dirt-bike</i> ¹² | minimoto | préf-rad | préfixation | trou comblé |
| <i>docudrama</i> | docufiction | rad-rad | composition | trou comblé |
| <i>drag and drop</i> | glisser-déposer | rad rad | composition | calque |
| <i>early booking</i> | réservation précoce | rad-suff rad | composition | calque |
| <i>fingerprinting</i> | empreinte numérique | rad rad-suff | composition | trou comblé |
| <i>flash-over</i> | embrasement généralisé éclair | rad-suff rad-suff rad | composition | trou comblé |
| <i>glossy</i> | laqué (matériel audiovisuel) | rad-suff | sémantique | traduction |
| <i>gold farmer</i> | collecteur d'or | rad-suff prep rad | synapsie | trou comblé |
| <i>iPod</i> | baladeur numérique | rad-suff rad-suff | composition | trou comblé |
| <i>junk food</i> | malbouffe (aliment ¹³) | préf-rad | sémantique | trou comblé |
| <i>morphing</i> | Recom. off. : morphose | rad-suff | suffixation | trou comblé |
| <i>naming</i> | Recom. off. : nommage | rad-suff | suffixation | traduction |
| <i>next gen</i> | nouvelle génération | rad rad-suff | composition | calque |
| <i>online</i> | en ligne | prép rad | conversion v | calque |
| <i>password</i> | mot de passe | rad prép rad | sémantique | calque |
| <i>pocket-bike</i> | minimoto | prép-rad | préfixation | trou comblé |
| <i>podcaster</i> | baladodiffuser | rad-rad | composition | trou comblé |
| <i>premarketing</i> | préplacement | préf-rad-suff | préfixation | trou comblé |
| <i>prime time</i> | première partie de soirée | rad rad prép rad-suff | synapsie | trou comblé |
| <i>product placement</i> | placement de produit | rad-suff prép rad | synapsie | calque |
| <i>resurfacing</i> | reliassage | préf-rad-suff | suffixation | trou comblé |
| <i>snacking</i> | Recom. off. : grignotage | rad-suff | sémantique | trou comblé |
| <i>social engineering</i> | ingénierie sociale | rad-suff rad | composition | calque |
| <i>spyware</i> | logiciel espion | rad rad | composition | trou comblé |

12 L'équivalent *minimoto* est employé aussi bien pour *dirt-bike* que pour *pocketbike*, alors qu'il s'agit de deux réalités différentes.

13 *Malbouffe* existe depuis longtemps comme concept en français. En revanche l'acception qui dénomme un aliment malsain est plus récente. Il peut s'agir d'une néologie sémantique par métonymie, ou d'un emprunt sémantique car les deux acceptions coexistent en anglais depuis longtemps. Comme dans les cas indécidables du paragraphe précédent, les deux explications ne sont pas exclusives l'une de l'autre.

| | | | | |
|---------------------|------------------------------|-------------------|-------------|-------------|
| <i>SUV</i> | VUS | let let let | siglaison | calque |
| <i>trojan</i> | cheval de Troie | rad prép rad | sémantique | trou comblé |
| <i>watermarking</i> | tatouage numérique | rad-suff rad-suff | composition | trou comblé |
| <i>web</i> | toile | rad | sémantique | traduction |
| <i>webcam</i> | Recom. off. : cybercaméra | rad-rad | composition | trou comblé |
| <i>weblog</i> | Recom. off. : bloc-notes | rad rad | sémantique | trou comblé |
| <i>webmaster</i> | webmestre | rad-rad | composition | adaptation |
| <i>wheeling</i> | (faire le) cabré | rad-suff | conversion | trou comblé |
| <i>worm</i> | ver (informatique) | rad | sémantique | traduction |

Tableau 2

2.1.3 Emprunts avec plusieurs équivalents

Il arrive fréquemment qu'un emprunt suscite, du moins dans un premier temps, plusieurs équivalents dans les langues cibles. Nous avons relevé (mais ce relevé ne prétend pas être exhaustif et recenser tous les équivalents en circulation pour les emprunts du corpus) douze emprunts et vingt-huit équivalents. Deux emprunts ont trois équivalents au moins et un en a au moins quatre. Le principe d'exposition des données (tableau 3) est identique à celui du tableau précédent à cette différence près que plusieurs lignes d'équivalents correspondent à un emprunt.

| Emprunt | Équivalents | Configuration | Matrice | Mode d'influence |
|-----------------------------------|---------------------------|--------------------------|-------------|------------------|
| <i>binge drinking</i> | cuite express | rad rad | composition | trou comblé |
| | biture express | rad rad | composition | trou comblé |
| | ivresse rapide | rad-suff rad | composition | trou comblé |
| | super biture | préf rad | préfixation | trou comblé |
| <i>bling bling</i> | nouveau riche | rad rad | sémantique | trou comblé |
| | tape-à-l'œil | rad prep rad | sémantique | trou comblé |
| <i>comfort food</i> ¹⁴ | aliment réconfortant | rad rad | composition | calque |
| | nourriture de consolation | rad-suff prép | synapsie | calque |
| | cuisine de réconfort | rad-suff rad prep rad | synapsie | calque |
| <i>crossover</i> | enjambement | préf-rad-suff | sémantique | trou comblé |
| | recombinaison | préf-rad-suff | suffixation | trou comblé |
| <i>flash mob</i> | foule éclair | rad rad | composition | trou comblé |
| | mobilisation éclair | rad-suff rad | composition | trou comblé |

14 Il est notable que la recherche d'une équivalence française aboutit à la création de trois calques, avec plusieurs synonymes pour chacun des deux termes anglais. Rien n'empêcherait l'apparition d'autres combinaisons : pour-quoi pas *aliment de réconfort*, *aliment de consolation*, *nourriture de réconfort*... ?

| | | | | |
|----------------------|--|----------------------|---------------------------|-------------|
| <i>phishing</i> | hameçonnage Recom. off. ¹⁵ : | rad-suff | suffixation | trou comblé |
| | filoutage | rad-suff | sémantique | trou comblé |
| <i>plug and play</i> | branché et joué Recom. off. : prêt à l'emploi | rad cc rad | composition | calque |
| | | rad prep rad | sémantique | trou comblé |
| <i>podcast</i> | comme podcasting ci-dessous | | | |
| <i>podcasting</i> | Rec. off. : diffusion par baladeur baladodiffusion | rad-suff prép rad | synapsie | trou comblé |
| | | rad-rad-suff | composition | trou comblé |
| <i>slum</i> | bidonville taudis | rad-rad | sémantique | traduction |
| | | rad | sémantique | traduction |
| <i>spam</i> | pourriel pollurriel pollupostage | rad-rad | mot-valise | trou comblé |
| | | rad-rad | composition ¹⁶ | trou comblé |
| | | rad-rad-suff | composition | trou comblé |
| <i>timeshifting</i> | décalage temporel lecture différée | rad-suff | composition | calque |
| | | rad-suff | | |
| | | rad-suff rad | composition | trou comblé |

Tableau 3

2.1.4 Équivalent seul, sans l'emprunt source :

comarqué (*cobranded*).

2.1.5 Récapitulatif des emprunts sans équivalent ou avec un ou plusieurs équivalents

| Sans équivalent | | Un équivalent | | Plusieurs équivalents | |
|-----------------|--------------------------|---------------|-------------|-----------------------|-------------|
| 47 | | 38 | | 12 | |
| assurés | déliçats ou complexes | emprunts | équivalents | emprunts | équivalents |
| 40 | 7 | 38 | 37 + 1 | 12 | 28 |

Tableau 4 : Récapitulatif des emprunts et de leur(s) équivalents éventuels¹⁷

Le total est de cent soixante-deux formes plus le cas unique d'équivalent sans l'emprunt correspondant. Les équivalents sont au nombre de soixante-sept et peuvent être classés de plusieurs points de vue.

¹⁵ *Hameçonnage* figure en fait comme deuxième choix possible dans les recommandations officielles.

¹⁶ *Composition* est un terme qui a été proposé par Fabienne Cusin-Berche (2003 : 34) pour dénommer des mots composés avec des éléments tronqués de deux mots, sans aucune partie commune, à la différence des mots-valises.

¹⁷ Dans ce tableau, on compte pour deux les deux emplois de *minimotos* (37 + 1) et on prend en compte les équivalents de *podcast* et de *podcasting*.

2.2 Typologie des « équivalents »

2.2.1 Les quatre ensembles

La première constatation qui s'impose est la grande disproportion entre les quatre ensembles, puisque le nombre d'unités de base (celles qui sont seules et celles qui figurent dans la première colonne) va de un (équivalent sans emprunt) à quarante-sept (pour les emprunts sans équivalents). Les deux ensembles qui comportent à la fois des emprunts et des équivalents en comptent trente-huit et douze. La multiplicité des équivalents constitue un indice de la nouveauté, avant une stabilisation dans le temps, une forme s'imposant au détriment des autres.

La faiblesse numérique des équivalents officiels sur ce corpus s'explique par le fait que nous avons traité essentiellement du lexique courant, tiré de la presse généraliste, et pas des termes venant de domaines de spécialité.

L'examen des équivalents créés par une matrice interne fait l'objet de plusieurs analyses, par matrices, par modes d'influence et par correspondances des matrices et des modes d'influence.

2.2.2 Par matrices des équivalents

On compte d'abord les matrices mises en action et la proportion qu'elles représentent sur les soixante-quatre néologismes par équivalence¹⁸ (tableau 5).

| Matrices | Emprunts avec un seul équivalent | Emprunts avec plusieurs équivalents | Total | % |
|-------------|----------------------------------|-------------------------------------|-------|------|
| Composition | 13 | 11 | 24 | 37,5 |
| Sémantique | 10 | 7 | 17 | 26,6 |
| Synapsie | 3 | 3 | 6 | 9,4 |
| Préfixation | 4 | 1 | 5 | 7,8 |
| Suffixation | 4 | 2 | 6 | 9,4 |
| Conversion | 2 | 0 | 2 | 3,1 |
| Mot valise | 1 | 1 | 2 | 3,1 |
| Compocation | 0 | 1 | 1 | 1,6 |
| Siglaision | 1 | 0 | 1 | 1,6 |
| Total | 38 | 26 | 64 | 100 |

Tableau 5

¹⁸ On ne prend pas en compte dans ces calculs le faux « équivalent » sans la source correspondante, et on ne prend en compte qu'une fois *diffusion par baladeur* et *baladodiffusion*.

2.2.3 Par modes d'influence

Le deuxième calcul prend en compte les modes d'influence de la langue source sur la langue cible. Rappelons que nous avons reconnu dans ce corpus quatre modes d'influence, par ordre progressif l'adaptation, la traduction, le calque et ce que nous appelons, faute de mieux, trou comblé (tableau 6). Il s'agit d'équivalents morphologiquement assez indépendants de la forme source.

| | Emprunts avec un seul équivalent | Emprunts avec plusieurs équivalents | Total | % |
|-------------|----------------------------------|-------------------------------------|-------|------|
| Trou comblé | 24 | 19 | 43 | 67,2 |
| Calque | 9 | 5 | 14 | 21,9 |
| Traduction | 4 | 2 | 6 | 9,4 |
| Adaptation | 1 | 0 | 1 | 1,5 |
| Total | 38 | 26 | 64 | 100 |

Tableau 6

On constate qu'il y a beaucoup d'équivalents assez indépendants, pour la forme, du mot qui a provoqué/justifié leur apparition mais, même au sein des trous comblés, il y a des degrés de proximité ou d'éloignement par rapport au mot source : on constate une nette recherche de se rattacher à la source comme gage d'efficacité (compréhension et diffusion) : *flash mob/mobilisation éclair*¹⁹.

2.2.4 Par un croisement des matrices et des modes d'influence

Le tableau suivant a l'ambition de croiser deux critères, les modes d'influence et les matrices. Il se lit dans les deux sens. Pour chaque mode d'influence, la première colonne indique le nombre d'items concernés, selon les matrices. La deuxième colonne, qui se lit verticalement, donne le pourcentage de chaque matrice au sein d'un mode d'influence. Les chiffres de la troisième colonne se lisent, horizontalement, avec ceux des cases correspondantes de la troisième colonne ; ils indiquent le pourcentage de chaque mode d'influence par rapport à la matrice.

¹⁹ Bien que *mobilisation* soit donné pour équivalent, notamment dans l'excellente explication de *Wikipédia*, il ne faut pas oublier que *mob* en anglais ne renvoie pas à *mobilisation* mais au latinisme *mobile vulgus* (étymologie bien attestée au XVII^e siècle). D'autre part, si *éclair* correspond bien à *flash* et pourrait conduire à une analyse par calque, l'emploi de *mob* pour *mobilisation* est inconnu en français. Une *mob* ne peut être qu'un vélomoteur (par antonomase de la marque *Mobylette*).

| Matrices | Modes d'influence | | | | | | | | | | | | Total Matrice | |
|-------------|---------------------|---------------------|------|---------------------|---------------------|------|---------------------|---------------------|------|---------------------|---------------------|-----|---------------|--------------------------------|
| | Trou comblé | | | Calque | | | Traduction | | | Adaptation | | | Matrice | % Matrice/ Total Matrice |
| | % Matrice/ MI | % MI/ Matrice | | % Matrice/ MI | % MI/ Matrice | | % Matrice/ MI | % MI/ Matrice | | % Matrice/ MI | % MI/ Matrice | | | |
| Composition | 16 | 37,2 | 66,7 | 7 | 50 | 29,2 | 0 | 0 | 0 | 1 | 100 | 4,1 | 24 | 37,5 |
| Sémantique | 11 | 25,6 | 64,7 | 1 | 7,1 | 5,9 | 5 | 83,3 | 29,5 | 0 | 0 | 0 | 17 | 26,6 |
| Préfixation | 5 | 11,6 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 7,8 |
| Synapsie | 3 | 7 | 50 | 3 | 21,5 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 9,4 |
| Suffixation | 5 | 11,6 | 83,3 | 0 | 0 | 0 | 1 | 16,7 | 16,7 | 0 | 0 | 0 | 6 | 9,4 |
| Conversion | 1 | 2,3 | 50 | 1 | 7,1 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3,1 |
| Mot valise | 1 | 2,3 | 50 | 1 | 7,1 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3,1 |
| Compocation | 1 | 2,3 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1,6 |
| Siglaision | 0 | 0 | 0 | 1 | 7,1 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1,6 |
| Total MI | 43 | 99,9 | | 14 | 99,9 | | 6 | 100 | | 1 | 100 | | 64 | 100 |

Tableau 7 : Croisement des matrices et des modes d'influence

Conclusion

L'analyse des résultats obtenus à partir de notre corpus confirme la nécessité d'opérer la distinction entre emprunts et créations sous influence, sous peine d'amalgamer des réalités différentes. Si tout emprunt résulte de l'influence d'une langue source sur une langue cible, toute influence externe exercée par une langue source sur une langue cible ne donne pas nécessairement un emprunt.

Une autre leçon de cette expérience réside dans la nécessité de prendre en compte la circulation relative des emprunts et de leur(s) équivalent(s) quand ils en ont.

L'extension d'enquêtes de ce type à d'autres langues permettrait d'observer, d'une part, les emprunts qui se diffusent internationalement, et, d'autre part, les capacités des langues à absorber des emprunts ou à créer des équivalents avec leurs propres ressources et quelles modalités elles mettent en œuvre pour cela – et ce ne sont sans doute pas les mêmes d'une langue à l'autre. Les leçons à tirer portent autant sur les systèmes linguistiques que sur l'évolution (voire la standardisation) de nos sociétés.

Bibliographie sommaire

- Cusin-Berche (F.), 2003 : *Les mots et leurs contextes*, Paris, Presses de la Sorbonne nouvelle.
- Goldis (A.), 1976 : « Calque linguistique » dans le cadre du contact entre deux langues apparentées : le français et le roumain », dans *Cahiers de lexicologie*, 28, p. 99-119.
- Jacquet-Pfau (C.), 2003 : « Du statut de l'emprunt en traitement automatique des langues », dans Sablayrolles (J.-F.), éd., *L'innovation lexicale*, Paris, Honoré Champion, p. 79-97.
- Jacquet-Pfau (C.) et Sablayrolles (J.-F.), 2008 : « Les emprunts : du repérage aux analyses. Diversité des objectifs et des traitements », dans *Neologica*, n° 2, p. 19-38.
- Humbley (J.), 2008 : « Emprunts, vrais et faux, dans le *Petit Robert 2007* », dans Pruvost (J.), éd., *Les journées des dictionnaires de Cergy : Dictionnaires et mots voyageurs. Les 40 ans du Petit Robert, de Paul Robert à Alain Rey*, Herblay, Éditions des Silves, p. 221-238, halshs.archives-ouvertes.fr/halshs-00275603/en/
- Pruvost (J.) et Sablayrolles (J.-F.), 2003 : *Les néologismes*, Paris, PUF (« Que sais-je ? »),
- Sablayrolles (J.-F.), 2000 : *La néologie en français contemporain*, Paris, Honoré Champion.
- Tournier (J.), 1985 : *Introduction descriptive à la lexicogénétique de l'anglais contemporain*, Paris, Champion-Slatkine.
- Tournier (J.), 1991 : *Précis de lexicologie anglaise*, Paris, Nathan.
- Winter-Froemel (E.), 2009 : « Les emprunts linguistiques : enjeux théoriques et perspectives nouvelles », dans *Neologica*, n° 3, p. 79-122.

Sandra de Caldas,
Laboratoire d'études romanes
et Département d'études des pays de langue portugaise,
Université Paris 8,
Saint-Denis,
France,
sdecaldas@yahoo.es.

Néologie terminologique et emprunt : résultat d'une interculturalité qui défie le temps. Analyse et traitement des emprunts lexicaux en français et en portugais dans un corpus officiel bilingue spécialisé dans le domaine de l'enseignement

Résumé

Le travail que nous nous proposons de divulguer est une approche théorique, analytique et comparative du phénomène de l'emprunt lexical dans deux langues romanes, français et portugais, à travers des textes officiels et spécialisés dans le milieu de l'enseignement.

Notre principale problématique est de comprendre les mécanismes de l'emprunt lexical du point de vue de la linguistique comparative. Pour cela, notre méthodologie de travail s'appuie sur les théories et les méthodes de la linguistique de corpus. Le corpus d'étude se compose des programmes scolaires du premier degré d'enseignement en France et au Portugal édités par les ministères de l'Éducation.

Mots-clés : néologie terminologique ; linguistique comparative ; emprunt lexical ; langue de l'enseignement ; dictionnarisation.

Introduction

De tout temps, le contact humain a permis le dialogue interculturel, le partage et les échanges. Ces interactions ont pour conséquences les grandes évolutions sociales, scientifiques et techniques, mais aussi linguistiques de nos sociétés. Ce vecteur de connaissances et de cultures s'enrichit des contacts et évolue avec le temps qui constitue un facteur incontournable. Parmi les principaux phénomènes linguistiques d'enrichissement des langues, on distingue l'emprunt comme le résultat d'un partage, d'une mixité ou encore d'un métissage culturel. Au Moyen-âge l'expansion arabo-musulmane véhiculait des savoirs scientifiques et techniques, les Grandes Découvertes ont laissé leurs traces, pendant la Renaissance, l'Italie rayonnait et a apporté son lot de désignations et, à l'ère de l'informatique, l'anglais

prédomine. Ces échanges sont « des témoins de l'histoire » comme le soulignait déjà Louis Deroy en 1956.

L'objectif principal de cette étude est d'observer le phénomène de l'emprunt lexical dans une langue de spécialité, dans deux langues romanes, afin de lever le voile sur certaines craintes que le concept d'emprunt peut susciter, notamment face à l'enjeu de la diversité linguistique. Notre principale problématique est de comprendre les mécanismes de l'emprunt lexical en situation spécialisée du point de vue de la linguistique comparative et plus particulièrement d'apprécier en quelles proportions cette langue spécialisée puise dans les autres systèmes linguistiques, de découvrir si cette dernière a souvent recours à des dénominations empruntées et finalement de voir comment les emprunts sont reçus et considérés en français et en portugais, afin de comprendre comment ces termes s'intègrent dans la langue et dans l'usage.

1 Approche théorique

1.1 Définitions de l'emprunt

Pour Louis Deroy, l'emprunt linguistique était « une forme d'expression qu'une communauté linguistique reçoit d'une autre communauté » (Deroy 1956 : 18). Louis Guilbert (1975 : 90) à son tour, définit l'emprunt comme « l'introduction à l'intérieur d'un système donné de segments linguistiques d'une structure phonologique, syntaxique et sémantique conformes à un autre système ». Au Québec, l'emprunt faisant l'objet d'une politique d'aménagement linguistique, Christiane Loubier en 2003, puis dans le texte adopté par l'Office québécois de la langue française en septembre 2007, définit l'emprunt linguistique comme « le procédé par lequel les utilisateurs d'une langue adoptent intégralement, ou partiellement, une unité ou un trait linguistique (phonologique, lexical, sémantique, syntaxique, etc.) d'une autre langue » (2003 : 21) et l'emprunt lexical comme « un emprunt linguistique par lequel les locuteurs d'une langue adoptent intégralement (forme et sens) ou partiellement (forme ou sens seulement) une unité lexicale d'une autre langue » (*idem*). Notons également que l'emprunt peut être interne ou externe; un emprunt peut provenir de langues étrangères, mais également avoir son origine à l'intérieur même d'un système linguistique.

Avec le temps, de très nombreux emprunts se fondent dans la masse lexicale d'une langue, notamment les plus anciens, difficiles à identifier sans une recherche approfondie. C'est pourquoi Christiane Loubier parle d'emprunt récent défini comme « un emprunt qui est dans l'usage effectif seulement depuis peu et qui n'est attesté que dans la documentation la plus actuelle » (*ibidem*) et d'emprunt ancien qu'elle qualifie « d'emprunt qui est dans l'usage effectif et qui est légitimé depuis longtemps par la société langagière » (2007 : 21).

Les unités échangées sont donc des emprunts malgré le temps qui passe, mais perdent leur caractère néologique au fil des années.

1.2 Emprunt et emprunt néologique

Dans le processus d'enrichissement des langues, le facteur temps joue donc un rôle important. Aujourd'hui, grâce aux nouvelles technologies et à la diffusion rapide des informations, certains linguistes, se proposent de fixer une durée limite accordée à la perception de nouveauté et d'autres, comme J.-F. Sablayrolles, affirment que « la durée de néologisme n'est pas uniforme et que certaines unités peuvent garder longtemps « un parfum de nouveauté », bien après leur première apparition, alors que d'autres perdent quasi instantanément ce caractère » (Sablayrolles 2002 : 101). I. Desmet, en 2002, s'interrogeait également sur la durée de perception d'un néologisme et déclarait que « pour qu'un terme puisse être considéré comme un néologisme, certains auteurs soutiennent qu'il doit exister depuis trois ans, d'autres voient cinq ans comme la limite » (2002 : 88) c'est justement « ce qui démontre bien la relativité de ce critère et du statut même des néologismes » (*idem*). Ceci ne fait que renforcer le fait que le processus d'intégration d'un nouveau mot est un défi du temps, qui traverse différentes phases et qui dépend de divers critères.

Mais alors quels sont les critères susceptibles de déterminer si une unité est néologique ou non ?

Pour définir un néologisme, I. Desmet parle de « critères de repérage des néologismes » (2002 : 86) et considère qu'ils peuvent être regroupés en quatre paramètres : la diachronie, la lexicographie, la nouveauté, l'instabilité (*idem*). En effet, le critère temps est primordial pour déterminer si une unité est néologique, mais également arbitraire puisqu'il varie en fonction des auteurs. Le critère lexicographique correspond à l'apparition ou non d'une unité dans les dictionnaires. J.-C. Boulanger, en 1979, considérait que « l'identification des néologismes terminologiques obéit à des raisons linguistiques. Pour cela, il faut distinguer les unités néologiques, c'est-à-dire celles qui n'ont pas encore été répertoriées dans les dictionnaires » (1979 : 54), mais il précise en 2009, après de longues années de recherches, que « rejoindre le dictionnaire n'enlève pas au mot sa qualité de néologisme. » (2009 : 374). J.-F. Sablayrolles, quant à lui, pense qu'on ne peut recourir systématiquement aux dictionnaires comme corpus d'exclusion pour la détection de néologismes, car ces derniers posent différents problèmes et sont incomplets et « prudents » dans l'intégration de nouvelles unités (Sablayrolles 2002). La néologie étant un concept « scalaire », J.-F. Sablayrolles prouve, après de nombreuses expériences en recherche néologique, qu'une unité lexicale peut être plus ou moins néologique et que le sentiment néologique fluctue « quantitativement et qualitativement » en fonction des chercheurs (Sablayrolles 2002). Les critères de nouveauté et d'instabilité quant à eux vont « de pair », selon I. Desmet (2002 : 88). Le premier, d'ordre psychologique, concerne le sentiment du locuteur face à un mot nouveau. Le second a donné lieu à de nombreuses études linguistiques d'ordre formel d'un point de vue morphologique, graphique, phonétique et sémantique puisque l'instabilité d'un néologisme est le fait qu'une « notion *a priori* nouvelle est désignée alternativement par deux ou plusieurs formations néologiques différentes » (Desmet 2002 : 86).

Ainsi, ces critères nous serviront de référence pour l'appréciation du caractère néologique des emprunts relevés.

1.3 Typologie de l'emprunt

Notre étude et nos recherches théoriques sur l'emprunt nous ont permis d'observer plusieurs formes d'emprunts, plusieurs phases d'évolution ainsi que plusieurs processus d'intégration dans la langue emprunteuse. Ainsi, en regroupant la théorie de l'école canadienne de C. Loubier (2003) et celle de J. Humbly (1974), nous avons élaboré la typologie suivante qui s'articule sur six niveaux de classification.

1. Tout d'abord, il s'agit de distinguer le genre d'emprunt, qui peut être lexical ou syntaxique.
2. Dans le cas d'un emprunt lexical, il faut ensuite distinguer la nature de l'unité : emprunt interne/emprunt externe, ce qui consiste à déceler un trait étranger à l'élément.
3. Puis, il faut isoler l'origine de l'unité. Dans le cas d'un emprunt interne, les échanges se font à l'intérieur d'une langue. Les unités pouvant passer d'un « domaine de spécialité » à un autre, de la langue générale à une langue de spécialité ou inversement. Dans le cas des emprunts externes, les emprunts peuvent provenir d'une langue étrangère ou savante.
4. Il est également possible de déterminer la catégorie de l'emprunt. Il faut donc distinguer l'emprunt sémantique (emprunt de sens), l'emprunt morphologique (forme et sens) ou l'emprunt formel (forme uniquement).
5. Lorsque l'on traite d'un emprunt sémantique, on peut isoler l'emprunt opaque ou calque linguistique de l'emprunt transparent ou calque morphologique. Lorsqu'il s'agit d'un emprunt morphosémantique, on peut distinguer l'emprunt intégral (sans aucune adaptation ou adaptation minimale) et l'emprunt hybride (avec différentes adaptations). Et finalement, dans le cas d'un emprunt formel, celui-ci peut être direct ou transformé.
6. Chaque type est encore classifiable en fonction des processus d'adaptation que peut subir un emprunt lors de son passage et de son intégration dans la langue emprunteuse.

| 1 | 2 | 3 | 4 | 5 | 6 | | | | |
|---------------------------------|---------|--------------------------------|------------------|-------------|---------------------|-------------------|---------------------|--|--|
| Genre | Nature | Origine | Catégorie | Type | Processus | | | | |
| L E X I C A L | EXTERNE | Savant | Sémantique | Opaque | traduction totale | | synonymie | | |
| | | | | Transparent | imitation de forme | ajout préposition | changement d'ordre | | |
| | | Langues Étrangères | Morphosémantique | Intégral | sans modification | apocope | adaptation minimale | | |
| | | | | Hybride | adaptation affixale | | semi-adaptation | | |
| | | | Formel | Direct | faux-emprunt | | | | |
| | | | | Transformé | | | | | |
| | INTERNE | Langue générale | | | | | | | |
| | | Langue spécialisée | | | | | | | |
| | | Patois Dialecte Géolecte | | | | | | | |

Tableau 1: Typologie de l'emprunt

2 Analyse et traitement

2.1 Méthodologie et corpus d'étude

Notre méthodologie de travail s'appuie sur les théories et les méthodes de la linguistique de corpus.

Tout d'abord, nous avons choisi de nous focaliser sur une langue spécialisée dans les sciences humaines : dans le milieu de l'enseignement du 1^{er} degré, domaine dans lequel s'emmêlent et s'échangent théories, pratiques et directives ministérielles et dont les écrits fleurissent à travers les siècles et varient selon les circonstances historiques et politiques. La jonction de ces concepts se trouve dans les plans d'étude qui dévoilent la pluridisciplinarité et la transdisciplinarité de ce domaine. Ces documents pédagogiques émanant des ministères de l'Éducation sont incontournables pour tout enseignant et fixent un ensemble d'objectifs à atteindre, d'activités à réaliser et de méthodes d'apprentissage. Ils sont la « référence de base » pour tous les enseignants. Ces textes officiels sont donc censés véhiculer une langue « respectueuse » des règles et des recommandations académiques.

La présente étude a été menée sur un corpus textuel informatisé et bilingue constitué du *Curriculo do ensino básico* (plans d'étude de l'enseignement du premier degré au Portugal) et du programme scolaire du premier degré d'enseignement en France

en vigueur depuis 2001. Afin de comparer les résultats obtenus, nous avons choisi des textes appliqués simultanément dans deux pays européens de langue romane. Notre corpus portugais représente un total de 118 063 mots et le corpus français un total de 100 583 mots.

2.2 Méthodologie de traitement du corpus

L'exploitation du corpus de travail s'est réalisée grâce à l'emploi de techniques de traitement semi-automatique des langues. L'utilisation de logiciels de traitement documentaire et statistique des corpus textuels, comme HYPERBASE ou le concordancier WORD SMITH, a permis d'isoler les termes spécifiques du domaine de l'enseignement ainsi que les emprunts lexicaux, mais aussi d'effectuer une exploration statistique afin de donner un aperçu quantitatif des emprunts dans le corpus. Grâce aux données recueillies, nous avons également procédé à un traitement qualitatif des unités empruntées, ce qui implique parallèlement une recherche diachronique de chaque terme pour déterminer les origines et les raisons de l'emprunt ainsi que les évolutions graphiques et sémantiques. Pour cela, l'utilisation de supports documentaires académiques s'est avérée indispensable, afin d'apprécier la réaction des institutions chargées des langues. Cette recherche permet également de mieux catégoriser chaque emprunt d'un point de vue synchronique.

3 Présentation et interprétation des résultats

Dans un premier temps, les textes qui constituent notre corpus d'étude présentent une certaine complexité terminologique. Un grand nombre d'unités terminologiques appartenant à divers « domaines » de spécialité ont été relevées. Ce qui confirme que cette langue de spécialité requiert des compétences dans différentes disciplines. Le domaine de l'enseignement a recours au vocabulaire de la psychologie, de la physiologie, de la sociologie et de toutes les disciplines enseignées, comme les mathématiques, les sciences, les arts, le sport, etc. D'une manière générale, la langue de l'enseignement est pluridisciplinaire et sa complexité réside dans sa diversité. Dans un deuxième temps, notre étude s'est focalisée sur les unités empruntées. Ainsi, nous avons pu les classer selon la typologie que nous proposons au point 1.3.

3.1 Classification des emprunts lexicaux

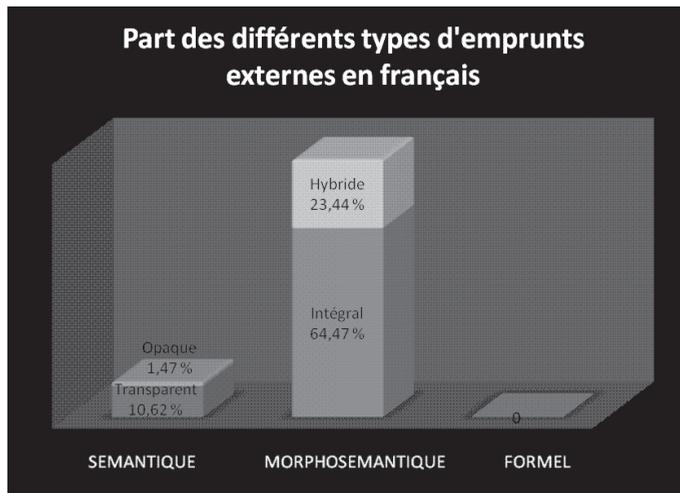
Tout d'abord, les emprunts lexicaux ont été traités de manière diachronique afin de déterminer leurs origines. Pour cet exercice, les supports documentaires de référence ont été les dictionnaires de langue à visée normative, soit les dictionnaires académiques, la huitième et neuvième édition du *Dictionnaire de l'Académie française* et le *Dicionário de língua portuguesa contemporânea da Academia das Ciências de Lisboa*.

En français, nous pouvons relever trois langues majoritairement prêteuses.

49,65 % des termes viennent de l'anglais (ex. : *discrimination, interaction, cédérom, dévédérom, scanner*), 21,37 % des termes viennent de l'italien (ex. : *bilan* et *attitude*) et finalement 20,02 % viennent de l'espagnol (ex. : *camarade*).

En portugais, les emprunts au français prédominent largement avec 72,94 % (ex. : *abordagem, atelier, dossier, motricidade*) et 10,78 % d'entre eux viennent de l'anglais (ex. : *nonsense, role playing, lettering, sketch, CD-Rom*).

Puis, les unités empruntées ont été catégorisées selon le type d'emprunt.

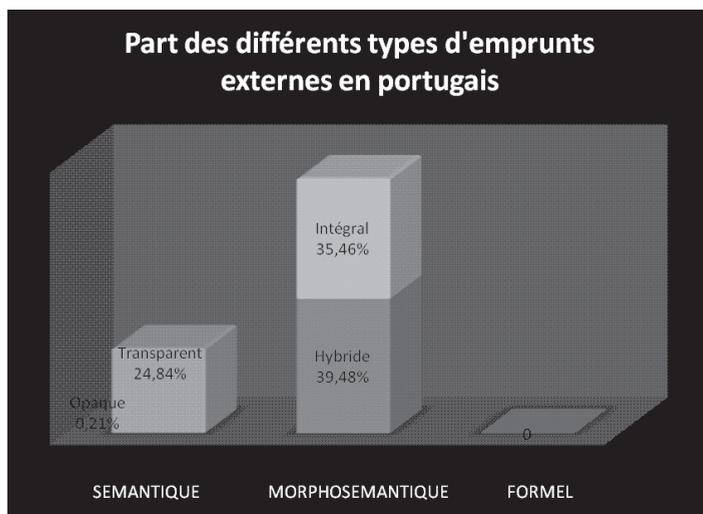


Graphique 1 : Part des différents types d'emprunts externes en français

Dans le corpus français, aucun emprunt formel n'est à relever.

En revanche, 12,09 % des emprunts externes sont des emprunts sémantiques avec une prédominance d'emprunts transparents. Le terme *discrimination*, par exemple, est un calque linguistique puisque le sens vient de l'anglais par le biais du philosophe Alexander Bain et a été attribué à une unité déjà existante en français.

Les emprunts morphosémantiques représentent 87,91 % des emprunts externes. 64,47 % sont des emprunts intégraux qui viennent essentiellement de l'anglais. Seuls 4,04 % ont été adaptés phonologiquement entraînant des répercussions graphiques minimales et 19,07 % ont été acceptés par apocope.



Graphique 2 : Part des différents types d'emprunts externes en portugais

Dans le corpus portugais, notons tout d'abord l'absence totale d'emprunt formel.

La majorité des emprunts externes sont des emprunts morphosémantiques et représentent 74,95 %.

Il est important de souligner que, parmi les emprunts intégraux, 65,75 % ont été adaptés phonologiquement avec des répercussions graphiques minimales, comme l'ajout d'un accent graphique. (ex. : *álbum* ou *dominó*). 8,02 % de ces termes ont été intégrés par apocope (ex. : *internet*, *site*, *skate*, *til* et *ténis*) et 26,23 % sont des emprunts intégraux qui n'ont subi aucune transformation.

Parmi les emprunts hybrides, 98,94 % sont des emprunts ayant subi des modifications de suffixes et proviennent essentiellement du français.

Les emprunts sémantiques représentent 25,05 % des emprunts externes en portugais, dont 24,84 % sont des emprunts transparents.

Finalement, en français, les emprunts à l'anglais sont peu adaptés alors que la langue portugaise adapte davantage les emprunts au français majoritaires dans le corpus d'étude portugais.

3.2 Dictionnarisation des emprunts lexicaux relevés

Un des quatre critères de repérage d'une unité lexicale néologique peut être le discutabile critère lexicographique. Le parcours des emprunts relevés dans le corpus d'étude a donc été étudié à travers différents dictionnaires de langue. Pour cela, un corpus lexicographique d'exclusion a été constitué à partir de dictionnaires normatifs et descriptifs, les plus courants en format papier ou consultables en ligne.

Le corpus d'exclusion français se compose des dictionnaires suivants : huitième édition du *Dictionnaire de l'Académie française* en ligne, neuvième édition du *Dictionnaire de l'Académie française* en ligne (publié depuis 1994 de A à Plébéien), *Littré* en ligne, *Trésor de la langue française (TLF)* en ligne, *Larousse* 1994, *Larousse* en ligne, *Grand Dictionnaire terminologique* en ligne, *Robert Micro de poche* 1983, *Robert Micro* 2006.

En français, la neuvième édition du dictionnaire à visée normative de l'Académie française ne dictionnarise que 87,41 % des emprunts relevés, ce pourcentage n'est pas totalement fiable puisque la neuvième édition n'est réactualisée que jusqu'à l'entrée « Plébéien ». Or, la huitième édition dictionnarise 88,15 % d'entre eux puisque *poney* et *zigzag* constituaient déjà une entrée dans cette version. Ainsi notons que les termes *puzzle*, *roller*, *scanner*, *site* et *slalomer* ne constituaient pas d'entrée dans la huitième édition, reste à vérifier s'ils seront dictionnarisés dans la neuvième édition non achevée. D'autres dictionnaires faisant autorité ne dictionnarisent qu'une partie des emprunts relevés, tels le *Littré* (83,33 %) et le *TLF* (84,82 %). Malgré les réactualisations des ouvrages en ligne, les dictionnaires de référence de la langue française n'ont pas encore laissé place à certains emprunts relevés dans cette étude. En revanche, 100 % des emprunts relevés sont dictionnarisés dans des ouvrages à visée descriptive comme le *Robert Micro* 2006. Notons qu'entre 1983 et 2006, le *Robert Micro* a intégré la totalité des emprunts relevés dans le corpus français. Les ouvrages dictionnaires de référence en français semblent plus « prudents » dans l'intégration de nouvelles unités alors que les ouvrages descriptifs de plus grande diffusion semblent mieux accompagner l'évolution de la langue.

Le corpus d'exclusion portugais, quant à lui, se compose des dictionnaires suivants : *Dicionário da Academia das Ciências de Lisboa* 2001, dictionnaires de la langue portugaise de *Porto Editora* 1982, 2006 et 2010, *Universal* 2010, *Priberam* en ligne, *Infopédia* en ligne basé sur les éditions de *Porto Editora*, base de données européenne *Iate*.

En portugais, une majorité d'emprunts relevés est dictionnarisée, et ce, même avant la parution des textes constituant le corpus. Ainsi, 96,80 % des emprunts relevés en portugais sont attestés par l'Académie des Sciences de Lisbonne. Cependant, le *Dictionnaire Porto Editora* 2010, fraîchement paru, intègre 97,08 % des emprunts relevés, tout comme la version 2006. Concernant les dictionnaires en ligne, notons que le site *Priberam* intègre davantage d'emprunts relevés dans notre corpus que le site *Infopédia*. Un certain nombre d'emprunts relevés ont été dictionnarisés dans l'édition de *Porto Editora* entre 1982 et 2010. Cependant, le dictionnaire *Universal* ne dictionnarise toujours pas certaines unités empruntées comme *design*, *jazz*, *kit*, *badminton* et *continuum*.

3.3 Instabilités et variation

D'une manière générale, l'étude de la dictionnarisation des emprunts permet de constater que les plans d'étude utilisent essentiellement des emprunts attestés. Cependant, quelques instabilités sont à noter.

En français, on peut observer deux phénomènes. Dans un premier temps, une différence de dictionnarisation est observable entre dictionnaires à visée normative et dictionnaires à visée descriptive. Certains emprunts ne sont pas intégrés dans les dictionnaires de référence de la langue française, alors qu'ils constituent des entrées dans les dictionnaires de grande diffusion et faisant l'objet de réactualisations régulières. Dans un deuxième temps, notons qu'il existe des instabilités graphiques entraînant parfois des situations de cohabitation. Prenons l'exemple de *cdérom* et *dévédérom* recommandés par l'Académie française. Ces termes cohabitent avec leurs formes vernaculaires *CD-ROM* et *DVD-Rom*, également dictionnarisés dans la langue française. Dans ces deux cas, les recommandations de l'Académie française semblent être respectées. D'autres emprunts vivent une situation de cohabitation avec la recommandation académique. Ainsi, *roller* n'est pas dictionnarisé par l'Académie française, qui recommande *patins à roulettes* et le *Grand Dictionnaire Terminologique* conseille *patins à roues alignées*. Même chose pour le terme *scanner*, l'Académie recommande *numériseur*. La forme étrangère de ces termes semble donc préférée aux recommandations académiques dans le corpus français.

Le terme « clip vidéo » souffre d'une instabilité de l'ordre des éléments. Le *TLF* définit le syntagme *vidéo clip* et le *Larousse* définit le terme « clip » et « vidéoclip ».

En portugais, dans un premier temps, une variation graphique engendrant une situation de cohabitation est observable. Certaines unités empruntées ont une double graphie : les emprunts *dossier* et *atelier* cohabitent avec une forme « portuguisée » *dossiê* et *ateliê* (les deux graphies apparaissant dans le dictionnaire de l'Académie des Sciences de Lisbonne). L'Académie recommande la forme « portuguisée » alors que, dans les plans d'étude, la forme d'origine est privilégiée.

Dans le cas du terme « vidéoclip », les recommandations ne sont pas suivies. Ce dernier est encore instable dans la langue, puisque d'autres formes cohabitent dans les ouvrages dictionnaires (*clipe de vídeo* ou *clipe et clip*).

D'autres termes apparaissent directement dans les plans d'étude portugais sous deux graphies différentes, une forme ayant subi une première adaptation à la langue réceptrice et une autre forme plus « portuguisée » : c'est le cas de *controle* (sans l'accent circonflexe du français) et *controlo* (avec modification de la lettre finale), *raquete* (réduction d'un *t* conformément à la règle du doublement de consonne en portugais) et *raqueta* (avec adaptation de la lettre finale) ou encore de *maquete* et *maqueta* (*idem*). Les deux graphies sont dictionnariées et apparaissent dans le

corpus portugais. Elles symbolisent deux phases successives d'adaptation d'une unité empruntée.

Dans un deuxième temps, un phénomène de variation typographique est également identifiable.

L'utilisation de signes typographiques est une marque d'instabilité et d'incertitude dans l'usage d'une unité. Ainsi, le terme *internet* ne semble pas encore tout à fait stable puisqu'il apparaît dans le corpus d'analyse tant avec une majuscule qu'avec une minuscule, le terme *lettering*, non dictionnarisé, apparaît en italique.

Tout comme *jazz* et *pop* ou encore *role playing* (non dictionnarisé) qui apparaît sous sa forme vernaculaire.

Le terme « *crol* » apparaît entre guillemets comme le terme « *nonsense* » et ne sont pas non plus dictionnarisés.

Dans un troisième temps, un phénomène de variation dictionnaire est à signaler.

Certains emprunts ne sont pas dictionnarisés, c'est le cas des termes *limógrafo*, *lettering*, *massmedia*. Le terme *crol* sous sa forme « portuguisée » apparaît dans notre corpus, mais est encore instable puisqu'il n'apparaît dans aucun dictionnaire de la langue portugaise. C'est la forme anglaise *crawl* qui constitue une entrée dans le dictionnaire de l'Académie des Sciences de Lisbonne. Ceci peut laisser croire que l'adaptation morphologique d'un emprunt en portugais se poursuit après l'attestation de ce dernier par l'Académie. *Badminton* apparaît dans le corpus sans accent alors que l'académie dictionnarise ce terme avec accent soit la forme « portuguisée ». Les deux graphies apparaissent dans certains dictionnaires, mais le terme « portuguisé » renvoie toujours à l'entrée sous sa forme non « portuguisée ».

Conclusion

Toutes ces observations permettent de dire que la langue de l'enseignement à travers les plans d'étude n'est pas une langue « envahie » par des emprunts néologiques. Cependant, des termes voyageurs plus ou moins récents sont utilisés dans les textes « officiels » émanant des ministères de l'Éducation. De « camarade » ou « discrimination » à « internet », « cédérom » et « roller », la présence de termes venus d'ailleurs, comme H. Walter (1997) les nomme, est indéniable.

En français, les emprunts relevés sont totalement intégrés, certains étant beaucoup plus anciens que d'autres. Les plus récents quant à eux semblent souffrir tout de même de quelques réticences.

En portugais, en revanche, nous pouvons considérer certains emprunts comme « néologiques » puisqu'ils semblent respecter les critères de repérage d'une unité néologique simultanément. En effet, des emprunts absolument pas dictionnarisés

ou partiellement dictionnarisés, présentant des instabilités graphiques et ressentis comme nouveaux du fait de leur traitement typographique ont été relevés.

Des situations d'instabilité graphique, de cohabitation et d'évolution post-dictionnarisation ou post-attestation sont observables en portugais. Or, en français, l'évolution post-dictionnarisation n'est pas constatée. De ce fait, il semble que l'Académie française atteste les emprunts avec plus de prudence. En revanche, deux graphies peuvent cohabiter, l'une attestée ou recommandée par l'Académie et l'autre par différents ouvrages dictionnaires à visée descriptive.

On a pu également relever des emprunts similaires en portugais et en français. Les termes « cédérom » et *CD-ROM*, « album » et *álbum*, « puzzle » et *puzzle* sont la preuve que les emprunts s'intègrent différemment dans les langues emprunteuses.

Face au faible taux d'emprunts externes relevés dans notre corpus bilingue, nous pouvons affirmer que le phénomène de l'emprunt ne semble pas affecter le vocabulaire de l'enseignement véhiculé par les plans d'étude, mais il lui permet d'accompagner les évolutions des techniques d'apprentissages axées aujourd'hui sur les nouvelles technologies.

Pour finir, toutes les instabilités relevées dans cette étude permettent évidemment de confirmer que le critère lexicographique n'est pas une garantie de « néologisme ». De plus, il semble évident que le locuteur reste maître quant à l'usage de la langue face aux recommandations officielles. Ceci peut laisser planer le doute sur le rôle et sur l'impact des institutions chargées de la langue et remettre en cause certaines normes. La langue est une plante vivante et dynamique qui varie et évolue avec le temps.

Bibliographie

- Alves (I. M.), 1995 : « Empréstimos nas línguas de especialidade: algumas considerações », dans *Ciência da Informação*, n° 3, vol. 24, Brasília, p. 319-321.
- Alves (I. M.), 2006 : « La dénomination par emprunt en portugais brésilien », dans *Neologica*, n° 1, LLI, Paris, p. 77-88.
- Boulanger (J.-C.), 1979 : « Néologie et Terminologie », dans *Néologie en marche*, série b, n° 4, Montréal, Office de la Langue française.
- Boulanger (J.-C.), 2009 : « Compte-rendu de *Neologica*. Revue internationale de néologie, n° 1 », dans *Meta: journal des traducteurs/Meta: Translators' Journal*, vol. 54, n° 2, juin 2009, p. 372-376.
- Deroy (L.), 1956 : *L'emprunt linguistique*, Paris, Les Belles Lettres.
- Desmet (I.), 2002 : « Néologie du portugais contemporain : une zone d'instabilité linguistique », dans *Actes de la journée « Instabilités linguistiques dans les langues romanes, travaux et documents »* n° 16, Presses Universitaires de Vincennes-Saint-Denis, p. 77-99.

- Desmet (I.), 2006 : « Variabilité et variation en terminologie et langues spécialisées : discours, textes et contextes », dans *Actes des Septièmes Journées Scientifiques du Réseau LTT : Mots, termes et contextes*, AUF, Paris, Édition des archives contemporaines, p. 235-247.
- Guilbert (L.), 1975 : *La créativité lexicale*, Paris, Larousse.
- Humbley (J.), 1974 : « Vers une typologie de l'emprunt linguistique », dans *Cahiers de lexicologie*, n° 25, Paris, Didier Érudition, p. 46-70.
- Humbley (J.), 2000 : « Quelques mécanismes de l'emprunt linguistique », dans *Actes de la journée en hommage à Louis Guespin, terminologie*, URA CNRS 1164, Université de Rouen, p. 71-77.
- Lino (T.), 2001 : « De la néologie à la lexicographie spécialisée d'apprentissage », dans *Cahiers de lexicologie 78-Hommage à Robert Galisson*, Paris, Honoré Champion, p. 139-145.
- Loubier (C.), 2003 : *Les emprunts : traitement en situation d'aménagement linguistique*, Québec, Langues et sociétés.
- Loubier (C.), 2007 : *Politique de l'emprunt linguistique*, Québec, Langues et sociétés.
- Sablayrolles (J.-F.), 2002 : « Fondements théoriques des difficultés pratiques du traitement des néologismes », dans *Revue française de linguistique appliquée*, VII-1, p. 97-111.
- Sablayrolles (J.-F.), 2003 : « Le sentiment néologique », dans *L'innovation lexicale*, J.-F. Sablayrolles dir., Paris, Honoré-Champion, p. 279-295.
- Sablayrolles (J.-F.), Cartier (E.) 2009 : « Néologismes, dictionnaires et informatique », dans *Cahiers de lexicologie*, Paris, Honoré Champion.
- Walter (H.), 1997 : *L'aventure des mots français venus d'ailleurs*, Paris, Robert Laffont.

Alicja Kacprzak,
Chaire de philologie romane,
Université de Łódź,
Łódź,
Pologne,
alicjakacprzak@wp.pl.

Diversité technolectale en diachronie : le cas de quelques termes médicaux français

Résumé

Lorsque les linguistes affirment que les langues évoluent, la plupart d'entre eux pensent à l'histoire de leur variante littéraire. Cependant, l'évolution concerne aussi d'autres variantes, qu'il s'agisse de dialectes, de sociolectes ou de technolectes. L'histoire de ces derniers est liée intimement à l'histoire du domaine qu'ils expriment. Mais l'analyse diachronique de quelques termes médicaux que nous proposons montre que l'intérêt d'une telle recherche dépasse les limites de simples constatations sur la terminologie ; en effet, elle fournit plusieurs informations sur la manière de concevoir le monde par les générations précédentes, et par conséquent sur la culture dite *partagée* des époques passées.

Mots-clés : diversité ; technolecte ; diachronie ; termes médicaux.

Introduction

Toute réflexion sur la diachronie des langues prend d'habitude en considération leur aspect standard, voire littéraire. Tel est notamment le cas de l'enseignement de la grammaire historique dans les facultés de philologies : des heures de cours sont consacrées à l'analyse de changements linguistiques relevés dans des œuvres littéraires, depuis la *Chanson de Roland* et Rabelais, jusqu'à Du Bellay et Hugo. Rien d'étonnant à ce qu'à la sortie des universités nos étudiants soient souvent convaincus que l'évolution ne concerne que la variante normée d'une langue, ou du moins que c'est seulement celle-ci qui est digne d'analyse historique. Une parenthèse s'ouvre, certes, en ce qui concerne la dialectologie : il est bien connu que l'existence de la variante standard de plusieurs langues modernes, dont le français, ne saurait être expliquée sans prendre en considération l'histoire de ses dialectes. Pour ce qui est par contre de la variation sociale, il est rare qu'un point de vue diachronique soit adopté dans ses analyses. Quelques dictionnaires d'argot ou quelques monographies rendent compte des états successifs du lexique non normatif, mais les études approfondies sur son évolution font défaut, sans doute aussi faute de corpus bien établi et fiable.

Dans cette situation, on peut s'étonner par contre que l'on attribue si peu d'importance aux analyses diachroniques des technolèctes et surtout de leurs lexiques, parfois si bien documentés à travers les siècles. Relativement rares sont en effet les travaux récents qui indiquent le contraire; parmi eux, il convient avant tout de citer l'article de Pascaline Dury et d'Aurélie Picton (2009), dressant un bilan de la recherche en terminologie historique et proposant en même temps un moyen pour remédier à cette situation.

Une raison de ce désintérêt pour la dimension temporelle des langues spécialisées est cependant certaine, c'est que les vocabulaires spécialisés sont presque uniquement envisagés dans leur utilité actuelle, qui consiste à dénommer efficacement, à bien servir la communication spécialisée. Les états précédents des appellations scientifiques ou techniques n'ont plus de valeur lorsque celles-ci sont envisagées selon un point de vue purement terminologique. Pourtant, ils constituent un vrai réservoir de connaissances sur les états antérieurs de la discipline en question, sur le niveau de spécialisation et, partant, sur la mentalité des professionnels, mais aussi des non-professionnels des époques antérieures. Dans le présent article, nous nous proposons d'en faire la démonstration en nous basant sur une analyse diachronique de quelques exemples appartenant au vocabulaire de la médecine – ce domaine particulièrement ancien et riche, qui depuis ses origines jusqu'à nos jours, a connu un essor sans pareil, tant quantitatif (nombre de faits médicaux découverts, recensés et décrits) que qualitatif (pertinence descriptive et diagnostique). Nous montrerons quels sont les cas les plus typiques de l'évolution des termes, en adoptant comme corpus une série d'écrits médicaux et généraux, parus entre le XVII^e et le XX^e siècle, afin d'illustrer le phénomène du changement terminologique, ses variantes et ses raisons.

1 Évolution des termes médicaux : des cas typiques

Une étude appliquée de l'évolution des notions et des termes choisis montre que quatre cas de figure particulièrement fréquents sont possibles : soit l'apparition d'un nouveau concept entraînant l'apparition d'un nouveau terme; soit, au contraire, la disparition d'un concept avec pour résultat la disparition du terme qui lui était attribué; soit encore la fluctuation de plusieurs termes autour du même concept à travers les siècles (le cas de la synonymie diachronique); soit enfin l'oscillation sémantique d'un seul et même terme dénommant deux ou plusieurs concepts à différents moments (le cas de la polysémie diachronique).

1.1 Un concept nouveau entraîne l'apparition d'un terme nouveau

En abordant l'analyse du vocabulaire spécialisé selon un point de vue diachronique, dans le cas du domaine presque éternel que constitue la médecine, on est confronté avant tout à un immense enrichissement terminologique au cours des siècles. C'est la découverte et l'évolution du savoir se trouvant à la base de la cristallisation de nouvelles notions qui entraînent un accroissement lexicogénique aussi important. Pour évaluer son ampleur, il suffit de comparer les chiffres : aux

2 800 noms de maladies répertoriées et décrites à la fin du XVIII^e siècle dans l'ouvrage de référence de l'époque, la *Nosologie méthodique* de Boissier de Sauvages de Lacroix (1772), correspondent 19 093 termes recensés dans la plus récente classification internationale des maladies, CIM-10, selon le répertoire de l'OMS. Deux voies typiques d'enrichissement terminologique se démarquent d'une manière particulière dans cette masse de termes créés, à savoir :

1.1.1 « Vagues » de termes comportant un élément nouveau commun

Chaque grande étape de l'évolution de la médecine apporte des vagues de termes liés à des faits médicaux qui viennent d'être découverts, souvent ces termes comportent un élément nouveau commun. Par exemple, l'époque de l'émergence de la neurologie en France (fin du XIX^e siècle), marquée entre autres par l'activité de deux grands neurologues, Jean-Martin Charcot et Joseph Babinski, a donné lieu à la création de plusieurs termes de ce domaine comportant le patronyme de l'un ou de l'autre : *ped de Charcot*, *syndrome de Charcot*, *signe de Babinski*, *épreuve de Babinski*, *syndrome de Babinski-Fröhlich*, *syndrome de Babinski-Nageotte*, *épreuve de la déviation angulaire Babinski-Weill*, etc. À son tour, une vague de découvertes en virologie dans les années soixante du XX^e siècle a eu pour résultat un nombre de termes comportant le morphème *vir-*, notamment : *viroplasma* (apparu en 1960), *viroïde* (en 1971), puis encore *virurie*, *virogène*, *virucide*, et autres.

1.1.2 « Séries » de termes construits sur un noyau hyperonymique

Une autre situation se présente dans le cas des appellations dans lesquelles des éléments nouveaux, apportant des précisions nécessaires, viennent s'ajouter au terme central, préexistant, ayant la valeur d'un hyperonyme. Prenons comme exemple le terme *anémie*. À la fin du XVIII^e siècle, Hélian (1771 : 5) l'évoque sans présenter aucun sous-classement du phénomène :

« Anémie : maladie qu'on peut regarder comme épuisement des vaisseaux sanguins et dont on n'a presque aucune mention. »

Presque cent ans plus tard, à côté du terme *anémie*, Jaccoud (1865 : 199-219) énumère ses variantes, dont *anémie des mineurs* et *anémie des altitudes*. E. Littré et Ch. Robin mentionnent bientôt après le terme *anémie épidémique des mineurs* et expliquent ainsi son origine : « anémie qui a régné épidémiquement parmi les ouvriers des mines de Chemnitz en 1777 » (1873 : 68). Le *Larousse médical illustré* (1924 : 60) ajoute encore quelques sous-types de la maladie : *anémie primitive*, *anémie infantile*, *anémie des vieillards*, *anémie tropicale*. Une nette différence par rapport à cette typologie est visible dans les ouvrages et dictionnaires médicaux contemporains, par exemple celui de Garnier et Delamare qui, dans l'édition de 1989, ne cite pas moins de 85 termes construits sur le noyau *anémie*, par ex. : *anémie hyperchrome*, *anémie hyperchrome hypersidérémique*, *anémie hyperchrome mégalocytaire*, *anémie hyperchromique*, *anémie hypochrome*, *anémie hypochrome hypsidérémique*, *anémie hypochromique*, etc. (Garnier et Delamare, 1989 : 40-43). Ces termes se présentant souvent sous l'aspect de définitions raccourcies et dont

la plus grande « technicité » par rapport aux termes plus anciens est évidente, reflètent la complexité grandissante des concepts à chaque étape de la découverte.

1.2 La disparition d'un concept entraîne la disparition du terme

La science médicale se développe aussi de telle manière que certains concepts, ancrés dans le savoir et la pratique séculaires, perdent de leur validité avec l'évolution de la discipline. Dans cette situation, la disparition du terme accompagne la disparition de la notion.

Si l'on considère les trois termes *plique mâle*, *plique femelle* et *plique polonaise*, cette triple distinction apparaît aussi bien dans le *Précis de la chirurgie pratique* en 1768 que dans la *Nosologie* de Boissier de Sauvages de Lacroix en 1772. Dans le premier ouvrage, le terme latin *plica polonica* est assimilé à *plique* ou *plie* tout court et l'on y mentionne *plique mâle* et *plique femelle* comme ses deux variantes.

« Nous distinguerons avec les auteurs deux sortes de plique. Nous appellerons une plique mâle et l'autre plique femelle. La plique mâle, plique en cordons, est la plus commune et se connoît en ce que les cheveux sont unis et collés, en manière de petits cordons. La plique femelle forme des pelotons de cheveux, indissolublement entrelacés, tantôt réunis, tantôt séparés, et qui forment une espèce de bonnet carré. Il coule souvent du sang des extrémités des poiles, ou des cheveux. » (Portal, 1768 : 332)

Pierre Boissier de Sauvages (1772), quant à lui, fidèle à son projet de systématiser les maladies, place les trois espèces au même niveau, en les subordonnant en même temps à un GENRE : *Trichoma*, qu'il place dans l'ORDRE : Anomalies et dans la CLASSE : Cachéxie.

L'ouvrage de F. L. de Lafontaine entièrement consacré au sujet de la plique indique :

« C'est en Pologne que la plique semble avoir fixé son siège principal. » (1808 : XI)

et explique plus loin :

« La plique polonaise est une maladie endémique en Pologne et dans les pays qui y confinent. Elle est occasionnée par une matière qui se jette d'une manière critique sur les cheveux, et les agglutine ensemble de telle sorte qu'il est absolument impossible de les peigner ou de les démêler. Ce principe morbifique ne se porte pas uniquement aux cheveux, il attaque aussi les ongles des pieds et des mains, principalement des personnes chauves. » (de Lafontaine, 1808 : 8)

Le classement qu'il en donne ne rejoint pas celui de Boissier, car la plique est divisée en plique vraie, plique fausse, plique vraie et fausse ; plique bénigne, plique pernicieuse ; plique simple, plique multiple, plique triple. Cette conception des variantes de la plique est valable encore au début du XIX^e siècle, quand Bégin (1823 : 472) présente sa nouvelle typologie de la plique : plique multiforme (*plica caput medusae*),

plique solitaire ou à queue, plique latérale, plique fusiforme, plique en masse globale. La prolifération des typologies et leur différenciation semblent constituer une preuve que la notion même de plique reste à l'époque sujette à une oscillation conceptuelle, ce qui affaiblit d'une manière évidente le statut des termes qui s'y appliquent. Il n'est point étonnant devant cette instabilité notionnelle que toutes les dénominations « satellites » de la plique disparaissent au cours du XIX^e siècle. Déjà au début du siècle, les auteurs de l'*Encyclopédie méthodique, médecine* en font ainsi état :

« PLIQUE : Les auteurs et les observateurs sont si peu d'accord sur cette maladie, qu'il est impossible d'en donner une distinction et de décider non seulement quel rang elle doit occuper dans un cadre morphologique, mais même si elle doit y trouver place. Selon les uns la plique n'est point une maladie [...] » (1827 : 149)

C'est donc à cette époque que, du moins dans la médecine officielle, la plique cesse d'être considérée comme une maladie. Un quart de siècle plus tard, Fabre parle seulement d'un « phénomène particulier se passant dans le système capillaire » (1850 : 417), et en 1924, le *Larousse médical illustré* l'évoque comme un « enchevêtrement inextricable des cheveux mêlés de parasites et de poussière » (1924 : 968), de même que le dictionnaire de Garnier et Delamare (1989 : 696), en lui ajoutant un synonyme *trichome*. La langue suit donc, dans ce cas, l'évolution du savoir scientifique. Ce qui semblait être une maladie grave aux XVII^e et XVIII^e siècles ne l'est plus, et même si le nom général de la pathologie reste en vigueur, ses variantes présumées disparaissent de la nosologie, ce qui fait par conséquent que des termes qui les dénommaient deviennent désuets et disparaissent à leur tour.

1.3 La fluctuation de termes autour du même concept

Dans certains cas, on observe aussi la fluctuation des termes se rapportant au même concept et l'on est ainsi confronté à une sorte de synonymie de termes en diachronie. Pour illustrer ce fait, il semble intéressant de suivre le parcours des trois termes *Pierre*, *calcul* et *caillou* dénommant un phénomène pathologique qui est défini en médecine comme « concrétion qui se forme parfois dans certains organes de l'homme ou des animaux ».

Selon le TLFi, le fonctionnement du mot *Pierre* au sens mentionné ci-dessus est attesté sous la forme écrite *piere* dans le *Lapidaire* de Marbode au XII^e siècle. Quant à l'appellation *calcul*, son apparition en 1546 est notée par le TLFi chez Ch. Estienne dans son *Discours des parties du corps* ; elle est en usage aussi chez Olivier de Serres qui conseille en 1603 dans le chapitre intitulé *Remèdes aux Maladies pour les Personnes* de son *Théâtre d'agriculture et mesnage des champs* :

« Pour la guérison de ce mal-ci, convient user de choses apéritives rompans le calcul, pour facilement uriner [...] » (2001 : 1350)

Mais l'emploi du terme *calcul* reste à l'époque si peu fréquent que le *Dictionnaire de l'Académie* le qualifie dans sa première édition (1694) de « vieux et hors d'usage » et

la dénomination de *Pierre* semble l'emporter tout au long du XVII^e siècle. Guy Patin (1601-1672) notamment, médecin et homme de lettres français, l'utilise à plusieurs reprises dans ses écrits. En 1634, il indique dans l'une de ses lettres :

« L'evesque d'Orléans est icy fort malade d'une pierre dans le rein, [...] »
(1907 : 67)

Dans une autre, écrite en 1640, il mentionne l'opération consistant à extraire des concrétions des organes atteints :

« l'anatomiste a quitté la reyne-mère à Londres, et est icy venue se faire tirer une grosse pierre de la vessie : l'opération en fut faite hier au matin [...] »
(Patin, 1846 : 68)

et ce, quel que soit le résultat de ces actes, comme le prouve la lettre du 1637 :

« Il y a quelques mois que M. Duret De Chevry, président des comptes, [...] mourut en cette ville, le troisieme jour après avoir esté taillé de la pierre [...] » (Patin, 1907 : 43)

Peu à peu cependant, ce terme commence à son tour à disparaître de l'usage. En 1771 encore, Hélian, dans son *Dictionnaire du diagnostic, ou de l'art de connoître les maladies, et de les distinguer exactement les unes des autres*, sous l'entrée *Pierre* (1771 : 343) explique :

« La Pierre dans les reins produit un genre de collique qu'on a appelé Néphrétique. »

Mais Pierre Boissier de Sauvages ne mentionne même pas ce terme dans sa *Nosologie méthodique*, en évoquant par contre celui du *calcul* :

« Mercatus prétend que le calcul a souvent percé les reins et cet accident a été suivi d'une ischurie incurable. » (1772, Tome x : 260)

L'influence connue de son œuvre sur l'aménagement du vocabulaire médical de l'époque semble être décisive aussi dans le cas analysé. Même si la dénomination *Pierre* (expliquée par *calcul*) apparaît encore en 1820 dans le *Dictionnaire des sciences médicales* sous l'entrée *maladie de la pierre* :

« Ce nom qui pourrait être appliqué à toute lésion morbifique dépendante de la formation ou de la présence d'un ou de plusieurs calculs [...] » (1820 : 427),

c'est désormais le terme *calcul* qui devient le plus utilisé. Ainsi, en 1870, le *Dictionnaire encyclopédique des sciences médicales* définit les *calculs* ainsi :

« concrétions solides qui se forment dans l'intérieurs des tissus. Quelques auteurs réservent plus spécialement le mot de calculs aux concrétions inorganiques. » (Dechambre, 1870 : 641)

En 1924, le *Larousse médical illustré* précise qu'il s'agit des :

« concrétions accidentelles qui se forment aux dépens des sels charriés dans les liquides de l'organisme [...]. Les dimensions des calculs varient d'un grain de sable fin à une masse de plusieurs kilogrammes. » (1924 : 187)

Et en 1965, l'*Encyclopédie médicale Quillet* donne cette description des complications possibles :

« Un calcul, après avoir franchi le cystique, peut arriver dans le cholédoque et exposera donc le malade à des complications en rapport avec cette nouvelle localisation. » (1965 : 156)

Force est de constater ainsi que pendant plus de deux siècles l'appellation *calcul* a monopolisé la désignation de la pathologie en question. Il est cependant intéressant de voir que le *Dictionnaire des termes de médecine* dans sa vingt-deuxième édition définit le « *calcul* » en faisant à nouveau recours au lexème *pierre* :

« Concrétion pierreuse, formée de substances organiques ou inorganiques qui prend parfois naissance dans les réservoirs glandulaires et dans les canaux extérieurs. » (Garnier, 1989 : 131)

Qui plus est, de nos jours, le terme *pierre* qui n'appartient pas au vocabulaire standard de la médecine est revenu par contre dans le parler populaire, comme cela est prouvé notamment par cette définition relevée sur un site web médical. Remarquons la mise entre guillemets du mot *pierre*, ce qui prouve encore son statut terminologique plutôt faible :

« Calcul de la vessie : “ pierre ” contenue dans la vessie constituée entre autres d'acide urique ou de phosphate de calcium ».

Les forums de santé dans Internet, cette *vox populi* de la Toile, offrent déjà un large éventail d'emplois du mot *pierre*, terme populaire pour *calcul*, qui constitue son équivalent standard, soit avec la conscience de ce fait comme dans :

« Les *calculs* rénaux appelés communément *pierres* au rein »,

soit directement, comme dans :

« J'ai une *pierre* de 6 mm de diamètre dans le *rein* droit. »

« Avez-vous aussi été traité pour une *pierre* au rein ? »

« J'aimerais savoir si vous avez une recette miracle pour éviter le retour des *pierres* sur les reins ».

Quant au troisième terme évoqué, *caillou* , sa carrière semble être la plus courte. Le TLFi n'atteste pas son emploi dans le sens étudié. On le trouve cependant déjà chez Alain (1936 : 98), qui écrivait le 2 janvier 1911 :

« [...] mélancolie: il y a quelque temps, je voyais un ami qui souffrait d'un caillou dans le rein, et qui était d'humeur assez sombre. Chacun sait que ce genre de maladie rend triste. »

Il semble être aussi assez répandu dans le langage populaire, comme on peut le constater sur Internet, aussi bien dans les définitions des sites consacrés à la santé :

« Également appelé *caillou* ou *Pierre*, le *calcul* rénal est une cristallisation qui se forme dans les reins ou les voies urinaires »

que sur de nombreux forums de discussions (citation avec l'orthographe originale) :

« bon jour j'ai des *caillou* dans la rein gauche qui me fait des douleurs atroces »
(*sic*)

« Bonne chance dans ta lutte contre ton *caillou* du rein gauche ».

Cet examen de l'histoire des trois appellations renvoyant au même phénomène permet de constater, après des siècles de concurrence entre *Pierre* et *calcul*, leur coexistence actuelle, avec cependant une nette opposition de type diastratique, *calcul* appartenant à la variante standard de la terminologie médicale (dans la variante scientifique c'est le terme *lithiase* qui est souvent employé par métonymie pour désigner le phénomène en question), *Pierre* à son tour, à la variante populaire. Il est intéressant de souligner que le terme *Pierre*, populaire aujourd'hui, fonctionne jusqu'au XVIII^e siècle comme un terme standard. *Calcul*, qui apparaît au XVI^e siècle, le supplante peu à peu. Quant au terme *caillou*, le recours à celui-ci semble découler d'un besoin d'une plus grande expressivité, les deux autres termes étant probablement devenus trop neutres pour un locuteur moyen. Aurait-il une valeur légèrement euphémistique ?

1.4 L'oscillation sémantique d'un seul terme dénommant plusieurs concepts

Il arrive pour des raisons diverses que le même terme soit employé au cours des siècles pour dénommer, à tour de rôle, différents concepts. Il s'agit alors d'une situation que l'on pourrait qualifier de polysémie diachronique. Une bonne illustration de ce cas est constituée par l'appellation *nyctalopie* qui désigne de nos jours un phénomène tout contraire à celui auquel elle se référerait au XVIII^e siècle. Ainsi, le *Dictionnaire de l'Académie* dans sa quatrième édition de 1762 définit la *nyctalopie* comme « maladie des yeux, qui fait qu'on n'y voit pas si bien le jour que la nuit ». De même, Hélian en 1771 décrit la *nyctalopie* comme « une maladie qui empêche de voir à quelque temps particulier de la journée où les autres voyent. Il n'y a aucun signe auquel on puisse reconnoître ces maladies; on n'en juge que sur la déposition des malades » (1771 : 285). Mais déjà cent ans plus tard, on trouve chez Jules Verne ce passage dans *Les Enfants du capitaine Grant* : « Il manœuvrait avec une surprenante habileté au milieu des ténèbres [...]. Sa *nyctalopie*, il est vrai, le servait fort, et ses yeux de chat lui permettaient de distinguer les moindres objets dans cette profonde obscurité » (Verne, 1868 : 181). Ce sens l'emporte aussi actuellement sur l'autre, ce que confirme le dictionnaire de Garnier et Delamare décrivant la *nyctalopie* comme : « faculté de voir la nuit. La *nyctalopie* (que l'on prête habituellement aux chats et aux rapaces nocturnes) est due à une augmentation des capacités des bâtonnets (cellules visuelles) à capter une faible quantité de lumière. Cette faculté se retrouve parfois chez certaines personnes » (1989 : 615). Notons cependant que le terme anglais *nyctalopie* garde toujours le sens ancien, à savoir « *day blindness* »,

confirmé par le *Dictionnaire Collins Français-Anglais*. Quelle qu'en soit la cause, le terme en question a subi une oscillation sémantique assez importante, car les deux acceptions présentent une valeur antonymique. Ce fait est certainement indicateur de la fragilité de certains termes moins fréquents, ce qui montre que leur implantation est un processus assez long et parfois imprévisible.

2 Intérêt des analyses diachroniques en terminologie

Une constatation évidente s'impose après notre brève analyse de quelques concepts et termes médicaux : ils sont d'une manière évidente tout aussi sujets à la variation diachronique que n'importe quelle notion ou n'importe quel mot du vocabulaire général. Ils évoluent, apparaissent, disparaissent et sont sujets à des changements d'ordre conceptuel et sémantique. Une étude attentive de ces changements fournit en même temps plusieurs informations non seulement sur les termes en question et leur évolution, mais aussi sur les états successifs de la discipline dont ils font partie et, de ce fait, sur la place qu'occupe cette discipline dans la culture générale de l'époque.

Dans le vocabulaire médical, l'évolution des termes illustre la manière spécifique de concevoir les concepts de telle manière que leur complexité s'accroît par rapport au concept de base, comme dans le cas de *anémie* → *anémies* (point 1.1.2). La prolifération terminologique qui en résulte montre que le découpage même de la réalité nosologique est un processus dynamique, en essor constant, qui est naturellement lié au progrès de la discipline.

L'évolution des termes reflète aussi la façon de percevoir des faits médicaux à différentes époques. Les dénominations anciennes sont le plus souvent sémantiquement transparentes, leur motivation s'appuie sur des valeurs directement perceptibles, comme notamment l'aspect du malade, ce qui est le cas, par exemple, de l'appellation *plique à queue* ou le groupe visé prétendu, comme dans le cas de la dénomination *anémie des jeunes filles*. Les termes plus récents se caractérisent par contre par leur « air » de plus en plus savant, ce qui constitue certainement l'un des signes de l'époque.

Des cas de fluctuations de termes rendent compte aussi de la façon dont évolue la conscience linguistique des spécialistes et des non-spécialistes, les uns et les autres faisant leurs choix terminologiques, en distinguant non seulement entre savant et populaire, mais aussi entre neutre et expressif, comme c'est le cas des quatre appellations quasi synonymiques *pierre/calcul/caillou/lithiase*.

Conclusion

L'histoire des termes apparaît donc comme à la fois complexe et intéressante. Les études ayant pour but la reconstruction de « la vie » des termes peuvent donner lieu à des considérations plus générales que celles concernant uniquement sur l'évolution

de la discipline : elles touchent aussi à l'évolution des mentalités, des cultures. Les analyses diachroniques en terminologie constituent donc une importante source de connaissances sur la manière de concevoir le monde par les générations précédentes et par conséquent sur la « culture partagée » des époques passées.

Bibliographie

- Alain, 1944, *Propos sur le bonheur*, 36^e éd., Paris, Gallimard.
- Bégin (L.-J.) *et al.*, 1823, *Dictionnaire des termes de médecine, chirurgie, art vétérinaire, pharmacie, histoire naturelle, botanique, physique, chimie, etc.*, Paris, Crevot, Béchet, Baillière.
- Boissier de Sauvages de Lacroix (F.), 1772, *Nosologie méthodique de distribution des maladies en classes, en genres et en espèces suivant l'esprit de Sydenham et la méthode des botanistes*, Tomes I-X, Lyon, Gouvion.
- Cabré (M.T.), 1998, *La terminologie. Théorie, méthode et applications*, Armand Colin, Les Presses de l'Université d'Ottawa, Ottawa.
- Dechambre (A.) *et al.*, 1886, *Dictionnaire encyclopédique des sciences médicales*, série 2, tome xxv, Paris, G. Masson, P. Asselin.
- Desmet (I.), « Variabilité et variation en terminologie et langues spécialisées : discours, texte et contexte » dans Blampain (D.), Thoiron (Ph.), Van Campenhoudt (M.), éd., *Mots, termes et contextes, Actes des Septièmes Journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction, Bruxelles, Belgique, 8, 9, 10 septembre 2005*, Paris, Éditions des archives contemporaines, p. 235-247.
- Dury (P.), Picton (A.), « Terminologie et diachronie : vers une réconciliation théorique et méthodologique », dans Humbley (J.), éd., *Terminologie, orientations actuelles, Revue Française de Linguistique Appliquée*, vol. XIV-2, décembre 2009, p. 17-31.
- Encyclopédie méthodique, médecine*, 1827, par une société de médecins, PHY-SEL, Paris, Vve Agasse.
- Fabre (F.A.H.) *et al.*, 1850, *Dictionnaire des dictionnaires de médecine français et étrangers ou Traité complet de médecine et de chirurgie pratiques, par une société de médecins*, Paris, Germer Baillière.
- Galtier-Boissière (J.) *et al.*, 1924, *Larousse médical illustré*, nouv. éd. entièrement refondue et augmentée par le D^r Burnier, Paris, Larousse.
- Garnier (M.) *et al.*, 1989, *Dictionnaire des termes de médecine*, 22^e édition, Paris, Maloine.
- Hélian (M.), 1771, *Dictionnaire du diagnostic, ou l'art de connoître les maladies, et de les distinguer exactement les unes des autres*, Paris, Vincent.
- Jaccoud (S.), 1865, *Nouveau dictionnaire de médecine de chirurgie pratique, illustré de figures intercalées dans le texte*, tome II, Paris, J.-B. Baillière.
- Kacprzak (A.), 2000, *Terminologie médicale française et polonaise – analyse formelle et sémantique*, Łódź, Wydawnictwo Uniwersytetu Łódzkiego.
- Kacprzak (A.), 2007, « Oscillation sémantique en tant qu'expression de la potentialité linguistique », dans Radimsky (J.), éd., *Synchronie dynamique du système linguistique, Écho des études romanes*, vol. III, n° 1/2, Ceske Budejovice, p. 137-142.
- Lafontaine, de (F.L.), 1808, *Le Traité de la plique polonaise*, trad. de l'allemand par A.J.L. Jourdan, Paris, Méquignon.

Littré (E.) et Robin (Ch.), 1873, *Dictionnaire de médecine, de chirurgie, de pharmacie, de l'art vétérinaire et des sciences qui s'y rapportent*, 13^e édition, Paris, J.-B. Baillière.

Patin (G.), 1846, *Lettres de Gui Patin, nouvelle édition augmentée de lettres inédites*, Tome troisième, Reveillé-Parise (J.-H), éd., Paris, Chez J.B. Baillière.

Patin (G.), 1907, *Lettres de Gui Patin, 1630-1672*, Triare (P), éd, Paris, H. Champion.

Portal (A.), 1768, *Précis de chirurgie pratique*, Paris, Vincent.

Serres de (O.), 2001, *Le théâtre d'agriculture et mesnage des champs: dans lequel est représenté tout ce qui est requis et nécessaire pour bien dresser, gouverner, enrichir et embellir la maison rustique*, Introduction Pierre Lieutaghi, Arles, Actes Sud.

Sournia (J.-C.), 1997, *Histoire de la médecine*, Paris, La Découverte/Poche.

Sites internet

www.e-sante.be

www.vulgaris-medical.com/encyclopedia

Ieda Maria Alves,
Departamento de Letras Clássicas e Vernáculas,
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo,
São Paulo,
Brésil,
iemalves@usp.br.

Néologie et société mondialisée : implications morphologiques

Résumé

Les néologismes créés dans une langue représentent les besoins quotidiens de dénomination et résultent fréquemment des événements politiques, historiques, sociaux, et aussi du développement scientifique et technologique acquis par la société qui parle cette langue. Dans cet exposé, nous avons pour but de mettre en relief les rapports que les néologismes vernaculaires établissent avec les faits sociaux qui déterminent leur création. Ces faits peuvent non seulement déterminer la création de nouvelles unités lexicales dans la langue parlée par cette société, mais aussi exercer de l'influence sur la structure morphologique de la langue, en déterminant la création de nouveaux formants ou en modifiant la fréquence d'usage de quelques-uns d'entre eux.

Mots-clés: néologie; néologisme; société; morphologie.

Introduction

Les unités lexicales néologiques créées dans une langue représentent les besoins quotidiens de dénomination et résultent très fréquemment des événements politiques, historiques, sociaux, entre autres, mais aussi du développement scientifique et technologique acquis par la société qui parle cette langue. Une grande partie des emprunts, ainsi que les termes des sciences et des techniques, représente cette face du néologisme.

Ces rapports entre société et vocabulaire ont été déjà soulignés par des néologues de différentes générations.

Matoré (1953 : 50), d'une manière explicite et même exagérée, est arrivé à classer la lexicologie en tant qu'une discipline sociologique : [...] *c'est en partant de l'étude du vocabulaire que nous essaierons d'expliquer une société. Aussi pourrions-nous définir la lexicologie comme une discipline sociologique utilisant le matériel linguistique que sont les mots.*

Les rapports entre néologie et société ont été aussi soulignés par Guilbert qui, dans *La créativité lexicale* (1975 : 31), a écrit : Mais le lexique ne consiste pas seulement dans le système de création lexicale, il débouche sur des unités de langage liées à l'univers des choses, aux modalités de la pensée, à tout le mouvement du monde et de la société.

Nous citons aussi Xu (2001 : 5) qui, à ce propos, souligne le rôle du néologisme comme élément qui reflète le mouvement de la société : *Le néologisme, étant la forme engagée des besoins nouveaux, constitue le signe du changement linguistique et des mutations sociales d'une époque donnée.*

Dans cet exposé, nous avons pour but de mettre en relief les rapports que les néologismes vernaculaires établissent avec les faits sociaux qui déterminent leur création et les implications morphologiques dues à ces rapports.

Nous prenons comme corpus les données des recherches de veille néologique que nous réalisons sur des matériaux extraits de la presse brésilienne (projet *Observatório de neologismos do português brasileiro contemporâneo (TermNeo)* — www.flch.usp.br/dlcv/neo). Il s'agit d'un projet de caractère lexicologique qui collecte des unités lexicales néologiques dans un corpus journalistique depuis janvier 1993. Dans une première phase, la veille néologique a pris comme base les journaux *Folha de S. Paulo* (FSP) et *O Globo* (G) et les revues *IstoÉ* (IE) et *Veja* (V), qui ont été analysés d'après un système d'échantillonnage (un journal ou une revue par semaine) : journal *O Globo* – premier dimanche du mois ; revue *IstoÉ* – deuxième semaine du mois ; journal *Folha de S. Paulo* – troisième dimanche du mois ; revue *Veja* – quatrième semaine du mois. En tant que corpus d'exclusion, ont été pris les dictionnaires de langue du portugais contemporain : Ferreira (1986) – corpus correspondant à la période de 1993-1999 ; Michaelis (1998) – corpus correspondant aux années 1999 et 2000 ; Ferreira (1999) – corpus correspondant à l'année 2000. Depuis 2006, un partenariat établi avec des chercheurs du *Núcleo interinstitucional de linguística computacional* (NILC), de l'Université de São Paulo – campus São Carlos – a permis le développement d'outils computationnels qui font la veille de néologismes de façon semi-automatique au moyen de lexiques informatisés qui servent de corpus d'exclusion. Ces outils, disponibles sur www.nilc.icmc.usp.br/~thiago/neologismo.html, ont permis l'augmentation du corpus d'analyse, qui prend maintenant en compte toutes les éditions en ligne des revues *IstoÉ*, *Veja* et *Época* (E) et deux éditions hebdomadaires des journaux *O Estado de S. Paulo* (ESP), *Folha de S. Paulo* et *O Globo*.

D'après ces données, les unités lexicales constituées d'éléments vernaculaires correspondent à 83 % des unités lexicales enregistrées et sont donc de loin les plus employées. Elles sont distribuées selon des modes de formation : dérivation préfixale (29 %) ; composition par subordination (17 %) ; composition par coordination (6 %) ; formation syntagmatique (12 %) ; dérivation suffixale (10 %) ; néologie sémantique (4 %) ; d'autres procédés (siglaison, troncation, mots-valises – 5 %). Les éléments empruntés correspondent à 17 % des données relevées.

1 Néologie, morphologie, société

Les unités lexicales néologiques peuvent être analysées sous différents points de vue (phonologie, morphologie, syntaxe, sémantique, texte, discours), étant donné que le lexique établit des rapports avec ces différents niveaux d'analyse linguistique, comme le font remarquer Mercè Llorente (2004 : 20) et Ieda Maria Alves (2009 : 1821). Nous soulignons, dans ce travail, les rapports que les unités lexicales néologiques établissent avec la morphologie, spécialement ceux qui sont déterminés par des faits sociaux.

1.1 Néologie et niveaux d'analyse linguistique

Les rapports établis entre la néologie et la phonologie peuvent être illustrés par le néologisme *anglo-sexônica*, où un changement de phonèmes – *sexônico* au lieu de *saxônico* – détermine la création d'une unité lexicale nouvelle :

- [1] *Os ingleses já tinham feito tentativas frustradas de aplacar a libido anglo-saxônica, quer dizer, anglo-sexônica, ao criarem o cricket, o futebol, o badmington.* (G, 06-12-98)

Au niveau de la syntaxe, des unités lexicales néologiques peuvent être créées par le recours à la conversion, selon lequel une unité – *chave* = clé –, appartenant à la classe des noms, peut être employée en tant qu'adjectif, en recevant un autre sens (important, principal) :

- [2] *Fêmea de borboleta azul grande, reintroduzida no Reino Unido há 25 anos, após ter sido extinta; proteção de formigueiros onde larva do inseto crescia foi chave para ressuscitar espécie.* (FSP, 16-06-09)

En ce qui concerne les rapports entre néologie et sémantique, nous citons le néologisme *adrenalina*, qui, outre le sens d'« hormone », est aussi employé dans le sens d'« enthousiasme, émotion » :

- [3] *Separe as montanhas e as noites frias de Campos do Jordão, tire os congestionamentos de carros e os ruídos das casas noturnas e adicione um pouco mais de adrenalina.* (FSP, 21-06-98)

Au niveau textuel, les unités lexicales néologiques établissent différents types de rapports.

Le sens d'une unité lexicale est parfois réitéré à l'aide d'un recours métalinguistique. Dans l'exemple suivant, le néologisme *megabanqueiro* est paraphrasé par « *maiores banqueiros* = les plus grands banquiers » :

- [4] *Outro megabanqueiro. Capitão do Unibanco e dono do Nacional, Pedro Salles virou um dos maiores banqueiros do país.* (V, 29-11-95)

Dans cet autre exemple, le suffixe *-ismo* (= *-isme*) attribuée à l'unité lexicale *blairismo*, dérivée du nom propre *Blair*, le sens d'une politique qui combine la croyance américaine dans le libéralisme avec le souhait européen de créer des institutions fraternelles :

- [5] *Chamemos de blairismo essa política que combina uma crença « americana » no liberalismo com um desejo « europeu » de criar instituições solidárias.* (V, 08-06-05)

1.1.1 Néologie et Société : implications morphologiques

Le rapport des unités lexicales néologiques avec les faits sociaux se révèle, d'abord, par la création vernaculaire de mots, qui traduit l'évolution culturelle et scientifique de la société, dans une période déterminée. Cependant, l'influence sociale dépasse la formation d'unités lexicales néologiques, car cette influence peut aussi déterminer des altérations dans le système morphologique d'une langue, en favorisant la création d'une structure morphologique dérivationnelle ou compositionnelle, ou en favorisant l'usage plus fréquent d'une structure déjà employée (Alves 2004).

Dans le cadre de la société brésilienne, un formant du portugais brésilien, l'adverbe *já* (« déjà »), a acquis un autre emploi grammatical dans les années quatre-vingt.

Depuis le coup d'État qui a eu lieu au Brésil en 1964, le peuple brésilien ne pouvait pas élire directement le président de la République. L'élection par voie indirecte, en octobre 1984, a été précédée d'un grand mouvement du peuple brésilien, appelé *a campanha das diretas-já* (« le mouvement des élections directes-déjà »). Les Brésiliens n'ont réussi à élire directement leur président qu'à partir de l'année 1989, mais depuis 1984, la composition de *directe* (adjectif en fonction substantivale) et de l'adverbe *déjà* se lexicalise et donne lieu à l'emploi de *diretas-já* :

- [6] *Ele acredita no restabelecimento do clima das « diretas-já » e do « fora-Collor » para « acabar com o projeto neoliberal do presidente Fernando Henrique ».* (FSP, 20-07-97)

Le caractère revendicatif exprimé par le néologisme *diretas-já* motive, depuis 1984, la formation d'autres néologismes avec *-já*, qui dénotent toujours une revendication, un souhait de changement. Nous citons *mudança-já*, c'est-à-dire, « un changement immédiat » :

- [7] *Afirmou que o governo não pode usar a necessidade de manter a moeda forte como argumento para adiar investimentos na área social: « O país quer mudanças-já ».* (FSP, 06-06-03)

Un autre exemple de l'influence de la société sur la structure morphologique est démontré par le préfixe portugais *sem-*, qui devient beaucoup plus fréquent dans la langue portugaise du Brésil à partir de la formation d'une unité lexicale nouvelle : *sem-terra* (= « sans-terre, travailleur rural qui n'a pas de terre »), attestée depuis les années cinquante. La création du *Movimento dos trabalhadores rurais sem-terra*

(mst), le Mouvement des travailleurs ruraux sans terre, a contribué à la formation de plusieurs néologismes avec *sans-*, qui marquent le « manque de quelque chose » :

[8] *Arc, o marciano, diz que quer entender quem são os sem-terra, os sem-teto, os sem-esperança, os sem-rumo...* (V, 25-08-99)

Le suffixe *-aço*, un vieux suffixe portugais qui forme des dérivés dénotatifs de « grandeur » et d'« intensité », a commencé à être très employé, en portugais brésilien, à cause des protestations qui ont eu lieu au Chili, à partir de l'année 1973, contre le gouvernement socialiste de Salvador Allende. En frappant sur des casseroles, les Chiliens ont créé le *cacerolazo* (mot créé avec le suffixe espagnol *azo*), qui a été calqué en portugais sous la forme *caçarolaço* (*caçarola* + *-aço*). Depuis cette création, beaucoup de formations nouvelles avec *aço* ont été remarquées dans le portugais du Brésil, en donnant même lieu à un emploi polysémique du suffixe, qui garde toujours le sens de « grandeur et d'intensité » (*lucraço* = « profit très élevé ») et a aussi acquis le sens de « très intense » (*apitaço* = « sifflement très intense ») :

[9] « *Se o fundo de pensão está em busca de lucratividade, deveria estar investindo em outras áreas* ». *Aproveitaram a oportunidade de trocar moedas podres nos leilões de privatização e agora saíram da jogada com moeda saudável. Conseguiram um lucraço.* (V, 01-05-96)

[10] *E, finalmente, assumiu ares de final de campeonato em dois episódios que mostram a semelhança entre políticos e cartolas: no mesmo dia em que os deputados da oposição promoveram um apitaço de meia hora no plenário para barrar a tramitação da emenda sobre a reforma administrativa,...* (IE, 14-05-97)

Un autre événement international a provoqué la création d'un nouveau formant affixal. Le formant *-gate*, qui, par la suite de l'épisode d'espionnage politique qui a eu lieu en juin 1972, à Washington, pendant la campagne électorale américaine, est souvent employé en fonction suffixale non seulement en anglais, mais aussi dans d'autres langues, y compris le portugais brésilien.

-gate n'est employé que dans des situations liées à un scandale ou à une dénonciation de corruption, comme le montre cet emploi relevé en 2005, qui révèle le détournement de l'argent public :

[11] *Desde o estouro do episódio já batizado de « Cuecagate », seu principal protagonista permanece mudo.* (V, 20-07-05)

D'autres formants sont employés dans le portugais brésilien, depuis le début des années quatre-vingt-dix, pour exprimer des aspects ou des besoins récents de la société contemporaine. Ils sont aussi employés en anglais et dans d'autres langues romanes.

Les progrès de l'informatique, qui ont surtout marqué la dernière décennie du xx^e siècle, ont déterminé l'usage croissant de *ciber-* (< *cybernetics*) et *e-* (< *eletronic*

mail), qui, venus de l'anglais américain, forment des néologismes associés aux innovations de l'informatique, tels que *cibercrime* et *e-golpe*:

- [12] « *Havia uma cultura do cibercrime na cidade* », afirma o delegado Cristiano Barbosa, responsável pela investigação. (V, 17-11-04)
- [13] *O e-golpe, como foi apelidado na Turquia, obrigou o primeiro-ministro, Recep Erdogan, a retirar a candidatura de Gul.* (E, 30-07-07)

Les luttes pour la préservation de la santé et de la planète Terre, ramenant sur le devant de la scène les anciens formants grecs *bio-* et *eco-*, marquent aussi les dernières années du siècle passé. Ces formants, autrefois restreints au domaine scientifique, se diffusent de plus en plus dans la langue générale et forment des néologismes qui traduisent les préoccupations contemporaines envers la santé et le monde où nous vivons. Nous citons les néologismes *biopirataria* et *ecotaxa*:

- [14] *O pesquisador é acusado de « crime contra a fauna e a flora » (o enquadramento legal de biopirataria), além de apropriação de dinheiro público.* (E, 29-06-07)
- [15] *O governo também pretende impor, a partir do início do ano que vem, uma « ecotaxa », que será cobrada de acordo com o tempo de permanência do turista.* (V, 29-08-01)

Un autre formant marque aussi les années contemporaines: *narco-*, qui résulte de la troncation de *narcótico* et représente les problèmes causés par la consommation et le commerce des drogues:

- [16] *É evidente que os países que apóiam a insensatez terrorista e os governos que toleram o narcoterrorismo não têm justificativa moral nem política para seus atos.* (V, 12-03-03)

Conclusion

Ces exemples démontrent que les faits qui ont lieu dans une société peuvent non seulement déterminer la création de nouvelles unités lexicales dans la langue parlée par cette société, mais peuvent aussi exercer de l'influence sur la structure morphologique de la langue, en déterminant la création de nouveaux formants affiaux ou en modifiant la fréquence d'usage de quelques-uns.

Ces formants, nouveaux ou réemployés, peuvent être restreints à une communauté linguistique, tels que *-já* et *sem-*. D'autres, comme *-aço* et *-gate*, peuvent être dûs à un fait survenu dans un autre pays. D'autres, comme *bio-*, *ciber-*, *e-*, *eco-* et *narco-* dépassent l'emploi dans une langue et peuvent être employés dans d'autres langues, en représentant la transculturalité des sociétés contemporaines.

Bibliographie

- Alves (I. M.), 2004 : « A unidade lexical neológica: do histórico-social ao morfológico », dans Isquierdo (A. N.) et Krieger (M. G.), éd., *As ciências do léxico*, vol. 2, p. 77-88.
- Alves (I. M.), 2009 : « Neologia e implicações textuais », dans *Anais do Congresso Internacional da Abralín — João Pessoa — 2009*, p. 1821-1825, <http://www.abralin.org/index.php>.
- Ferreira (A. B. de H.), 1986 : *Novo dicionário da língua portuguesa*, 2e édit., Rio de Janeiro, Nova Fronteira.
- Ferreira (A. B. de H.), 1999 : *Novo Aurélio século XXI: o dicionário da língua portuguesa*, 3e édit., Rio de Janeiro, Nova Fronteira.
- Guilbert (L.), 1975 : *La créativité lexicale*, Paris, Larousse.
- Lorente (M.), 2004 : « A lexicologia como ponto de encontro entre a gramática e a semântica », dans Isquierdo (A. N.) et Krieger (M. G.), éd., *As ciências do léxico*, vol. 2, p. 19-30.
- Matoré (G.), 1953 : *La méthode en lexicologie*, Paris, Didier.
- Michaelis, 1998 : *Moderno dicionário da língua portuguesa* : São Paulo, Melhoramentos.
- Xu (Z.), 2001 : *Le néologisme et ses implications sociales*, Paris, L'Harmattan.

*Mirabela Navlea,
Amalia Todiraşcu,*
Équipe LiLPa-Linguistique, Langues, Parole,
Université de Strasbourg,
Strasbourg,
France,
mirabela_abe@yahoo.com,
todiras@unistra.fr.

Ressources linguistiques pour un outil de traduction automatique statistique factorisée français-roumain

Résumé

Cet article présente un projet de recherche qui se concentre sur la résolution des problèmes de la traduction automatique (TA) français-roumain. À cet effet, nous utilisons un système de TA statistique factorisée s'appuyant sur des corpus parallèles annotés et alignés. Nous exploitons des corpus alignés au niveau lexical. Comme l'alignement lexical génère un nombre important d'erreurs pour différentes paires de langues, nous avons relevé ces erreurs, leur correction permettant d'améliorer l'alignement. Dans cet article, nous présentons les résultats de l'analyse des corpus alignés français-roumain et les solutions proposées.

Mots-clés: système factorisé; corpus alignés; alignement lexical.

Introduction

Un premier objectif de notre projet est d'étudier l'influence des informations linguistiques (morphologiques, syntaxiques) sur la qualité de la TA statistique. Notre deuxième objectif est de définir des ressources linguistiques adaptées pour un outil de TA statistique factorisée français-roumain.

Notre étude se situe dans la lignée du projet *SEE-ERA.net* (Tufiş *et al.* 2008, Académie roumaine de Bucarest) dont l'objectif était le développement d'outils de TA statistique factorisée pour des langues slaves et balkaniques (bulgare, grec, roumain, serbe, slovène) de et vers l'anglais. Pour la traduction automatique, ce projet utilise des corpus parallèles alignés au niveau lexical et annotés (étiquettes morphosyntaxiques, lemmes, constituants syntaxiques partiels).

Les systèmes de TA ont une place de plus en plus importante sur le Web, vu que le nombre des documents multilingues y augmente vertigineusement. Nous voyons le besoin de traduire des documents en temps réel, de comprendre l'utilisateur et de lui répondre rapidement, tout en traitant une quantité importante de documents.

Les outils de TA répondent également à des besoins précis en termes de création et de gestion de contenu multilingue, de création d'outils de recherche performants.

Cependant, la plupart des logiciels actuels de TA traduisent de et vers l'anglais. Nous constatons que très peu de systèmes de TA traitent le roumain. Pour le roumain-anglais, nous pouvons citer les travaux de Marcu et Munteanu (2005), Irimia (2008), Ceașu (2009) et *Google Translate*¹. Le système *Google Translate* est le seul outil qui propose la paire de langues français-roumain.

Malgré les avancées techniques actuelles, les systèmes de TA commettent un nombre important d'erreurs dues au manque de ressources linguistiques performantes (dictionnaires, grammaires) pour différentes langues. La création de telles ressources est une tâche difficile conditionnée par de nombreux phénomènes linguistiques qui se produisent dans une langue donnée (ambiguïtés, manque d'équivalents de traduction d'une langue à l'autre, phraséologie). Grass (2009) relève treize types d'erreurs générées par les logiciels de TA, parmi lesquelles la polysémie et l'homonymie, l'ambiguïté (syntaxique, référentielle), les néologismes, etc. Pour pallier ces problèmes, les outils de TA doivent mettre en place des ressources linguistiques complexes (dictionnaires de synonymes, de collocations, bases terminologiques ou de connaissances) ou les résoudre partiellement à l'aide des techniques statistiques utilisant des corpus alignés.

Dans le domaine de la TA coexistent deux approches : l'experte et l'empirique. L'approche experte se base sur des règles linguistiques complexes (lexicales, d'analyse syntaxique, etc.) coûteuses en temps et en ressources humaines et matérielles. Les systèmes de TA statistique factorisée (*EuroMatrix*, 2009) utilisant des corpus parallèles annotés et alignés obtiennent des résultats similaires aux méthodes expertes et s'avèrent moins coûteux.

Les systèmes factorisés (Koehn et Hoang 2007, Tufiş *et al.* 2008) représentent une extension des systèmes statistiques standard à base de séquences (Koehn, Och et Marcu 2003) par l'exploitation de facteurs linguistiques associés au mot (lemmes, étiquettes morphosyntaxiques, etc.). Ces systèmes s'avèrent performants puisqu'ils offrent la possibilité de moduler les ressources linguistiques utilisées. Koehn et Hoang (2007) montrent que l'intégration d'informations morphologiques riches améliore nettement les résultats du système de traduction. Avramidis et Koehn (2008) obtiennent des résultats prometteurs en exploitant l'information syntaxique dans la phrase source, qui permet l'extraction de l'information linguistique demandée par la phrase cible, à partir des langues pauvres vers des langues riches morphologiquement. Les résultats des systèmes factorisés dépendent des paires de langues considérées et aussi du sens de traduction. Ainsi, nous évaluons les facteurs linguistiques pertinents et nous appliquons plusieurs combinaisons afin de les intégrer dans le processus de traduction automatique français-roumain pour obtenir des résultats optimaux.

1 translate.google.com/translate_t?hl=fr#.

Dans les sections suivantes, nous présentons le fonctionnement du système de TA statistique factorisée que nous adaptons et la méthodologie adoptée pour développer notre système factorisé français-roumain. Nous décrivons notre corpus parallèle de travail, les étapes de prétraitement et d'alignement du corpus. Nous analysons les erreurs d'alignement lexical et nous proposons des règles heuristiques pour améliorer la qualité de l'alignement lexical utilisé par le système.

1 L'architecture du système de traduction automatique statistique factorisée

Nous présentons l'architecture du système de TA statistique (Koehn 2008), constitué par trois composantes (Figure 1) :

- un modèle de traduction obtenu par une analyse statistique d'un corpus parallèle aligné ;
- un modèle de langue pour la langue cible obtenu par une analyse statistique d'un corpus monolingue ;
- un décodeur algorithmique qui fournit la plus probable traduction de la phrase source, parmi toutes les traductions possibles, à partir des modèles de traduction et de langue construits auparavant.

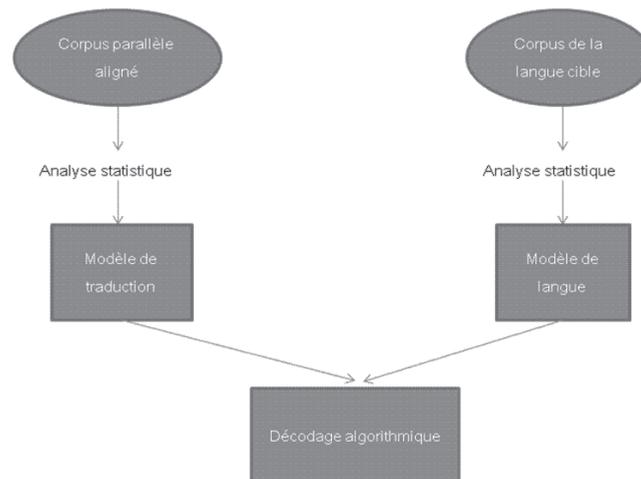


Figure 1 : L'architecture du système de TA statistique

Le système factorisé exploite des corpus alignés qui sont préalablement annotés, à la différence du modèle à base de séquences qui n'intègre pas d'informations linguistiques. Nous utilisons le décodeur *MOSES* (Koehn *et al.* 2007) pour lequel nous développons des ressources linguistiques appropriées. *MOSES* étend la traduction de séquences par l'utilisation des informations linguistiques supplémentaires associées aux mots et a l'avantage d'être un système *open source*.

Un système factorisé met en correspondance les facteurs linguistiques associés aux mots et ensuite exploite ces facteurs pour générer les formes fléchies correctes des mots. Nous illustrons ces étapes sur un exemple de traduction de l'anglais vers le roumain (Ceașu 2009), pour une configuration possible du système *MOSES*.

Premièrement, le système associe les lemmes qui sont suivis par les deux premiers caractères des étiquettes morphosyntaxiques :

treaty^{Nc} → *tratat*^{Nc}

Cette technique permet la désambiguïsation des lemmes. Par exemple, le lemme *tratat* (= *traité*) du roumain peut être nom commun, adjectif participial ou participe passé du verbe *a traTA* (= *traiter*).

Deuxièmement, le système associe les étiquettes des parties du discours :

NN → *NSN, NSRY, NSOY, NPN, NPRY, NPOY*

L'étiquette de partie du discours *NN* (nom, sans déterminant) du corpus anglais est mise en correspondance avec toutes les étiquettes possibles concernant le nom dans le corpus roumain : nom singulier ou pluriel (*NS, NP*), non déterminé (*NSN* ou *NPN*), déterminé (*Y*), dans les cas nominatif-accusatif (*R*) ou génitif-datif (*O*).

Dans un troisième temps, le système associe les étiquettes morphosyntaxiques des unités lexicales considérées :

Nc-s-n → *Ncms-n, Ncmsry, Ncmsoy, Ncfp-n, Ncfpry, Ncfpoy*

L'étiquette morphosyntaxique *Nc-s-n* (nom commun, singulier, sans déterminant) du corpus anglais est mise en correspondance avec toutes les étiquettes possibles du nom commun dans le corpus roumain : nom commun, masculin, singulier, non déterminé ou déterminé dans les cas nominatif-accusatif ou génitif-datif ; nom commun, féminin, pluriel, non déterminé ou déterminé dans les cas nominatif-accusatif ou génitif-datif.

Finalement, le système génère toutes les formes fléchies du mot cible à partir des résultats de sortie des étapes antérieures :

tratat^{Nc} | *NSN* | *Ncms-n* | → *tratat*

tratat^{Nc} | *NSRY* | *Ncmsry* | → *tratatul...*

Ainsi, le système génère toutes les formes fléchies du mot cible (*tratat, tratatul, etc.*) à partir du lemme, des étiquettes des parties du discours et des étiquettes morphosyntaxiques de sortie. Par conséquent, cette méthode permet la traduction des mots supplémentaires inconnus initialement dans les corpus d'apprentissage. Par exemple, si le mot d'entrée *treaty* est connu dans le système et la forme *treaties* est inconnue, le système peut générer la forme cible *tratate* (= *traités*) à partir du

lemme connu et des descriptions linguistiques de sortie *tratat^Nc|NPN|Ncfp-n*. Cette technique est utile surtout pour les langues riches morphologiquement, comme le français et le roumain.

Nous présentons notre méthodologie pour construire un système factorisé français-roumain dans la section suivante.

2 La méthodologie

Dans notre projet, nous développons un outil de TA statistique factorisée implémenté initialement pour l'anglais et le roumain² (Ceaşu 2009). Ce système utilise un corpus parallèle aligné aux niveaux propositionnel et lexical et une combinaison de facteurs linguistiques, tels que :

- les formes des mots ;
- les lemmes ;
- les étiquettes morphosyntaxiques (le jeu d'étiquettes du projet *Multext* (Ide et Véronis 1994)) ;
- les *chunks* ;
- les collocations.

Comme le système n'a pas encore été adapté pour la paire de langues français-roumain, notre étude se concentre sur l'identification des paramètres linguistiques à intégrer dans les modèles de langue et de traduction que nous construisons avec *MOSES* (Koehn *et al.* 2007).

Afin de construire les ressources nécessaires à notre système, nous adoptons la méthodologie suivante :

- la constitution d'un ensemble de corpus parallèles et monolingues ;
- le prétraitement (tokenisation, lemmatisation, étiquetage, *chunking*) et l'alignement propositionnel et lexical des corpus ;
- la correction des erreurs d'alignement dans les corpus utilisés pour l'apprentissage ;
- la construction de modèles de langue et de traduction français-roumain à partir des corpus ;
- la configuration du système avec les facteurs linguistiques les plus pertinents ;
- l'analyse linguistique des erreurs de traduction, la reconfiguration du système et la reprise de l'étape 4, afin d'obtenir un résultat optimal.

2 www.racai.ro/webservices/FactoredTranslation.aspx.

Dans la section suivante, nous présentons le corpus parallèle de travail et les prétraitements effectués, ainsi que les étapes et les erreurs d'alignement lexical français-roumain.

3 L'analyse des erreurs d'alignement

3.1 Le corpus parallèle

Le corpus de travail est *JRC-Acquis* (Steinberger *et al.* 2006). Ce corpus multilingue est composé des normes européennes des années 1950-2008. Il est disponible dans les 22 langues officielles de l'UE et comprend entre 30 et 60 millions de mots par langue. Nous en avons extrait un ensemble de documents communs aux deux langues étudiées.

Nous avons passé le corpus parallèle français-roumain par une étape de prétraitement et notamment de tokenisation, d'étiquetage morphosyntaxique et de lemmatisation. Pour cela, nous avons utilisé l'outil *TTL* (*Tokenizing, Tagging and Lemmatizing free running texts*)³ (Ion 2007) disponible pour l'anglais, le français et le roumain. Cet outil réalise aussi une analyse syntaxique partielle du corpus, le *chunking*. Ainsi, le corpus dispose d'une annotation par *chunks* (groupe nominal, prépositionnel).

L'outil fournit le fichier de sortie en format XML (Figure 2). Dans cet exemple, les balises *w* encadrent chaque unité lexicale. Les attributs *lemma* et *ana* des balises *w* représentent le lemme et respectivement la description morphosyntaxique des unités lexicales. Les étiquettes *ana* sont en format *Multext* (Ide et Véronis 1994). L'outil a reconnu aussi les groupes marqués par l'attribut *chunk*.

```
<seg lang = "fr"><s id = "ttlfr.3">
<w lemma = "voir" ana = 'VmPs.s"> VU </w>
<w lemma = "le" ana = "Da-fs" chunk = "Np#1"> Ia </w>
<w lemma = "proposition" ana = "Ncfs" chunk = "Np#1"> proposition
</w>
<w lemma = "de" ana = "Spd" chunk = "Pp#1" > de </w>
<w lemma = "le" ana = "Da-fs" chunk = "Pp#1, Np#2"> la </w>
<w lemma = "commission" ana = "Ncfs" chunk = "Pp#1,
Np#2"> Commisslon </w>
<c> ; </c>
</s> </seg>
```

Figure 2 : Un exemple de phrase traitée avec TTL

3 ws.racai.ro/ttlws.wsdl.

De plus, le corpus roumain dispose d'une annotation au niveau de dépendances (relation Prédicat – Complément). L'information sur les dépendances est extraite vers le roumain à partir des corpus alignés roumain-anglais (Ion *et al.* 2006). Nous utilisons l'analyseur *Cordial* pour le corpus français afin d'obtenir la même information sur les dépendances. Cette information sera ensuite utilisée pour améliorer la qualité de l'alignement lexical du corpus parallèle.

Nous avons corrigé les erreurs d'étiquetage et/ou de lemmatisation sur une partie du corpus (environ 500 000 mots pour chaque langue) pour augmenter la précision des résultats de l'alignement lexical exploitant des informations linguistiques.

3.2 L'alignement propositionnel

L'alignement propositionnel du corpus parallèle est effectué par le biais d'un classifieur de type vecteur-support *SVM* (Ceauşu, Ştefănescu et Tufiş 2006). Cet outil fonctionne en deux étapes indépendamment des langues de travail. L'outil réalise un alignement basé sur la longueur des propositions et sur les équivalents de traduction. L'aligneur fournit le fichier de sortie en format XML.

Nous avons gardé seulement les alignements de type 1-1, c'est-à-dire une proposition est traduite par une seule proposition. Nous avons procédé de cette manière puisque la recherche des unités à aligner ne dépasse pas les frontières de la phrase pour la plupart des outils actuels d'alignement lexical.

Après l'alignement propositionnel du corpus parallèle, nous avons procédé à l'alignement lexical présenté dans la section suivante.

3.3 L'alignement lexical

Par manque de ressources disponibles pour effectuer directement un alignement lexical français – roumain, nous avons utilisé l'anglais comme langue pivot (Tufiş et Koeva 2007). En effet, nous avons généré l'alignement lexical français-roumain à partir des corpus alignés anglais-roumain et anglais-français. L'opération de dérivation s'appuie sur l'idée que si les alignements initiaux sont corrects, l'alignement final le sera aussi. L'alignement lexical anglais-français est réalisé avec l'outil statistique *GIZA ++* (Och et Ney 2003) et corrigé manuellement, tandis que l'alignement anglais-roumain est obtenu avec l'aligneur *COWAL* (*Combined Word Aligner*) (Tufiş *et al.* 2006) exploitant des informations linguistiques. *COWAL*, disponible pour l'anglais et le roumain, est une combinaison de deux aligneurs – *YAWA* et *MEBA* – qui ne génèrent pas les mêmes erreurs, donc ces méthodes peuvent être combinées pour obtenir de meilleurs résultats.

YAWA utilise un lexique d'équivalents de traduction obtenus automatiquement (*TREQ*, Tufiş 2002, *GIZA ++*) et filtrés par plusieurs heuristiques, comme les cognats (mots similaires orthographiquement, ayant un sens commun) ou les correspondances entre les catégories lexicales (par exemple, les verbes peuvent être traduits par des noms, mais pas par des conjonctions). *YAWA* comprend trois étapes.

Premièrement, l'outil aligne les mots-contenu (noms, verbes). Deuxièmement, l'aligneur effectue l'alignement des *chunks* et des mots à l'intérieur de ces *chunks* par un module dépendant de la paire de langues considérées. Finalement, les mots qui sont restés non alignés précédemment sont alignés par des règles heuristiques.

MEBA utilise un algorithme itératif pour générer les alignements pas à pas. La première itération effectue les alignements ancrés par l'utilisation des cognats, des dates et des nombres. L'itération suivante aligne les mots-contenu dans la proximité immédiate des alignements ancrés. Dans l'optique de l'aligneur *MEBA*, un alignement est caractérisé par un ensemble de propriétés divisées en propriétés indépendantes de contexte (équivalence de traduction, correspondance entre les catégories lexicales, cognats) et propriétés dépendantes de contexte (collocations, localité – le degré de regroupement des alignements). *MEBA* fait aussi une autre distinction entre ces propriétés qui peuvent être bidirectionnelles (équivalence de traduction, correspondance entre les catégories lexicales) ou non directionnelles (cognats, localité, collocations).

L'évaluation de *COWAL* montre des résultats performants en termes de précision (86,99 %), de complétude (79,91 %) et de *F-measure* (83,30 %) (Tufiş *et al.* 2006).

Dans la section suivante, nous présentons les résultats de l'analyse des corpus alignés français-roumain, ainsi que les règles heuristiques proposées pour l'alignement lexical.

3.4 Les erreurs d'alignement

Nous avons relevé une série d'erreurs systématiques d'alignement lexical dont nous présentons ici quelques exemples. Ces erreurs sont dues principalement aux différences morphosyntaxiques entre les deux langues étudiées.

Le français et le roumain sont des langues romanes avec une morphologie flexionnelle riche qui présentent des similarités quant à la structure syntaxique assez proche, mais nous constatons aussi l'existence des différences morphosyntaxiques importantes entre les deux langues. Ainsi, en roumain, les nominaux présentent des formes casuelles (nominatif, accusatif, génitif, datif, vocatif) et le genre neutre. Au niveau des clitiques, le déterminant défini du roumain est toujours enclitique et fusionne avec le nom ou avec l'adjectif antéposé, tandis qu'en français il est toujours proclitique. D'autres différences concernent les constructions spécifiques de chaque langue (les structures relatives) et les morphèmes supplémentaires d'une langue à l'autre (l'expression de la relation de possession). Vu ces considérations, nous avons défini des schémas morphosyntaxiques qui seront utilisés comme heuristiques par *COWAL* afin de pouvoir effectuer directement l'alignement pour cette paire de langues, sans passer par l'anglais.

Au niveau du lexique, nous constatons que beaucoup de néologismes du roumain viennent du français. Ainsi, nous exploitons des listes riches de cognats pour filtrer

les équivalents de traduction. Nous sommes en train de constituer ces listes pour l'alignement lexical français-roumain.

Nous avons repéré des erreurs concernant l'alignement des collocations. Les collocations sont des expressions polylexicales dont les mots entretiennent une relation lexico-syntaxique (Todiraşcu *et al.* 2008) et peuvent avoir comme équivalents de traduction aussi bien des collocations, qu'une seule unité lexicale. Par exemple, une collocation verbo-nominale comme *avoir le droit* a comme équivalent roumain la collocation *a avea dreptul*, mais *procéder à l'examen* se traduit par le verbe *a examina* (= *examiner*). Par ailleurs, une collocation verbo-nominale comme *mettre en application* peut être traduite par sa variante nominalisée *punerea în aplicare* (= *la mise en application*). Ainsi, nous utilisons un dictionnaire multilingue de collocations munies de leurs propriétés morphosyntaxiques contextuelles (Todiraşcu *et al.* 2008) pour aligner en bloc ce type d'expressions. Les collocations deviennent ainsi des ancres fortes pour le squelette de l'alignement.

Nous avons analysé les erreurs d'alignement lexical du corpus *JRC-Acquis*, dues aux différences morphosyntaxiques entre le français et le roumain. Sur cette base, nous avons défini 27 schémas morphosyntaxiques parallèles pouvant être utilisés comme règles heuristiques par l'aligneur lexical (*cf.* section 3.3). Ces ressources sont utilisées principalement pour l'alignement des *chunks* et des mots à l'intérieur de chaque *chunk*. Les règles définies peuvent résoudre ainsi les erreurs qui apparaissent principalement au niveau des déterminants et des prépositions, des constructions spécifiques de chaque langue, des morphèmes supplémentaires (marques de cas). Nous présentons ci-dessous quelques exemples d'erreurs d'alignement lexical repérées dans le corpus *JRC-Acquis*.

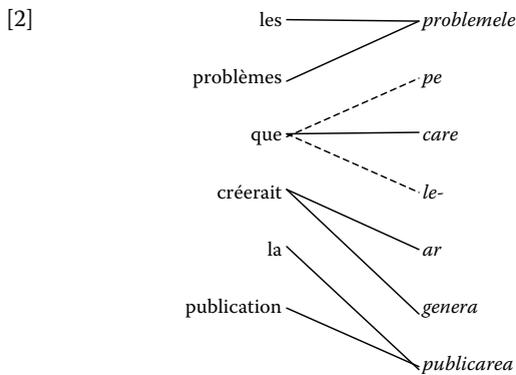
Une des erreurs fréquentes concerne l'expression de la relation de possession. En effet, les marqueurs de cette relation ne sont pas alignés. Pour exprimer la possession, le français utilise la préposition *de*, tandis qu'en roumain une des possibilités est l'emploi d'un morphème supplémentaire marquant le cas génitif *al, a, ai, ale* suivi d'un nom en génitif :

| | | |
|-------------|-------|------------|
| [1] l' | _____ | instituția |
| institution | _____ | competentă |
| compétente | _____ | a |
| d' | _____ | unui |
| un | _____ | stat |
| état | _____ | membu |
| membre | _____ | |

Nous donnons ci-dessous les schémas morphosyntaxiques définis pouvant résoudre ce type d'erreur :

| | |
|----|--|
| fr | déterminant défini + N + ADJ + <i>de</i> + déterminant indéfini + N |
| ro | N (déterminant défini) + ADJ + <i>al</i> <i>a</i> <i>ai</i> <i>ale</i> + déterminant indéfini (forme de génitif) + N génitif |

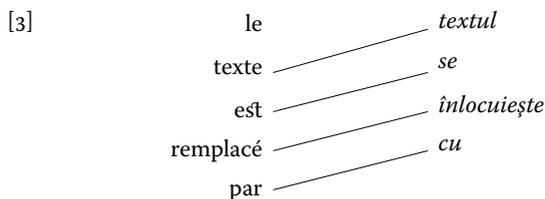
Une autre catégorie d'erreurs fréquentes concerne les structures relatives. Dans les subordonnées relatives du roumain, le COD est doublement exprimé par le pronom relatif *care* en accusatif, toujours précédé par la préposition *pe*, et les formes non accentuées du pronom personnel *îl*, *-l*, *o*, *îi*, *-i*, *le*. De même, le COI s'exprime par le relatif *care* en datif et les formes non accentuées du pronom personnel *îi*, *-i*, *le*. Nous donnons ci-dessous un exemple d'erreur où le pronom relatif COD *que* du français n'est pas aligné avec la préposition *pe* et le pronom personnel *le* du roumain :



Les schémas morphosyntaxiques définis pour ce type de structures sont les suivants :

| | |
|----|--|
| fr | N + <i>que</i> + V |
| ro | N + <i>pe</i> + <i>care</i> (accusatif) + <i>îl</i> <i>-l</i> <i>o</i> <i>îi</i> <i>-i</i> <i>le</i> + V |

Le français utilise souvent les constructions passives, tandis que le roumain préfère les constructions pronominales dans les corpus étudiés. Dans l'exemple ci-dessous, le verbe auxiliaire *être* qui apparaît en français n'est pas aligné avec le pronom réfléchi *se* du roumain :



Nous donnons ci-dessous les schémas morphosyntaxiques définis pour corriger ce problème :

| | |
|----|--|
| fr | N + <i>être</i> + V participe passé + préposition <i>par</i> |
| ro | N + <i>se</i> + V + préposition <i>cu</i> |

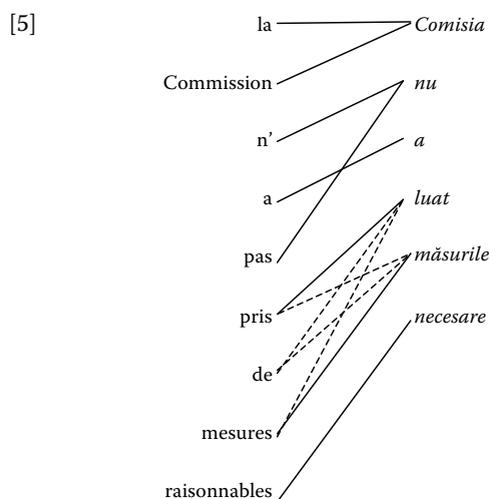
L'infinitif français se traduit souvent en roumain par le subjonctif. Dans ce cas, la particule *să* du subjonctif roumain n'est pas alignée avec l'infinitif français :



Les schémas morphosyntaxiques ci-dessous permettent d'éviter ce type d'erreur :

| | |
|----|-------------------|
| fr | V + V infinitif |
| ro | V + <i>să</i> + V |

Nous donnons ci-dessous un exemple d'erreur d'alignement lexical concernant les collocations qui ne sont pas alignées en bloc. Par conséquent, des unités lexicales appartenant aux collocations restent non alignées. Dans cet exemple, le déterminant indéfini *de* appartenant à la collocation du français *ne pas prendre de mesures*, à la forme négative, reste non aligné :



Nous résolvons ce type d'erreur d'alignement lexical par l'utilisation des propriétés morphosyntaxiques représentées dans un dictionnaire multilingue de collocations comprenant aussi le français et le roumain (Todiraşcu *et al.* 2008).

Conclusion

Nous adaptons un système de TA factorisée pour le français et le roumain, à l'aide des corpus alignés et annotés avec des informations linguistiques. L'alignement lexical du corpus parallèle requiert une précision élevée pour assurer la qualité des modèles de traduction. Nous étudions les erreurs existantes dans le corpus aligné pour améliorer la qualité de l'alignement. Les erreurs sont dues aux différences morphosyntaxiques entre les langues étudiées et nous proposons des règles heuristiques pour les corriger.

Nous combinons des méthodes statistiques avec des méthodes utilisant ces informations linguistiques (COWAL, Tufiş *et al.* 2006) pour réaliser un alignement français-roumain. Ainsi, nous utilisons des équivalents de traduction obtenus à l'aide de GIZA ++ (Och et Ney 2003) que nous filtrons en utilisant des ressources linguistiques (listes de cognats, dictionnaire multilingue de collocations) et des mesures statistiques appropriées (correspondance de catégories lexicales, localité). Nous utilisons également la base de règles heuristiques morphosyntaxiques définies suite à l'analyse des corpus alignés.

Dans le futur, nous étudierons aussi la possibilité d'intégrer des informations syntaxiques comme ressources pour l'alignement lexical.

Nous appliquerons cette méthode d'alignement sur d'autres corpus parallèles constitués dans le cadre de notre travail. Nous comparerons également les résultats d'alignement lexical obtenus avec les résultats fournis par GIZA ++ afin de choisir la meilleure méthode d'alignement.

Nous construirons des modèles de traduction à partir de l'alignement lexical français-roumain. Nous procéderons ensuite à des évaluations successives de ces modèles et à l'optimisation de leurs résultats.

Bibliographie

- Avramidis (E.) et Koehn (P.), 2008 : « Enriching Morphologically Poor Languages for Statistical Machine Translation » dans *Proceedings of ACL-08: HLT*, Columbus, June 2008, p. 763-770.
- Ceaşu (A.), Ştefănescu (D.) et Tufiş (D.), 2006 : « Acquis Communautaire Sentence Alignment using Support Vector Machines », dans *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, May 2006, p. 2134-2137.
- Ceaşu (A.), 2009 : *Tehnici de traducere automată si aplicabilitatea lor limbii române ca limbă sursă*, Thèse de doctorat, Académie Roumaine, Bucarest, avril 2009, 123 p.
- EuroMatrix*, 2009 : *The EuroMatrix Project: Statistical and Hybrid Machine Translation Between All European Languages*, dernière mise à jour le 28 février 2009, www.euromatrix.net/.
- GIZA ++, 2003 : *Training of statistical translation models*, dernière mise à jour le 30 septembre 2003, www.fjoch.com/GIZA++.html.

- Grass (T.), 2009 : « À quoi sert encore la traduction automatique? », dans *Les Cahiers du GEPE, Outils de traduction - outils du traducteur?*, n° 3, Strasbourg, décembre 2009, 14 p.
- Ide (N.) et Véronis (J.), 1994 : « Multext (multilingual tools and corpora) », dans *Proceedings of the 15th CoLing*, Kyoto, August 5-9, p. 90-96.
- Ion (R.), 2007 : *Metode de dezambiguizare semantică automată. Aplicații pentru limbile engleză și română*, Thèse de doctorat, Académie Roumaine, Bucarest, mai 2007, 148 p.
- Ion (R.), Ceaușu (A.) et Tufiş (D.), 2006 : « Dependency-Based Phrase Alignment », dans *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, May 2006, p. 1290-1293.
- Irimia (E.), 2008 : « Experimente de Traducere Automată Bazată pe Exemple », *Atelierul de Lucru Resurse Lingvistice Românești și Instrumente pentru Prelucrarea Limbii Române*, Iași, 19-21 novembre 2008, p. 131-140.
- Koehn (Ph.), Och (F. J.) et Marcu (D.), 2003 : « Statistical Phrase-Based Translation », dans *Proceedings of HLT-NAACL 2003 Main Papers*, Edmonton, May-June 2003, p. 48-54.
- Koehn (Ph.) et Hoang (H.), 2007 : « Factored translation models », dans *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CONLL)*, Prague, June 2007, p. 868-876.
- Koehn (Ph.), Hoang (H.), Birch (A.), Callison-Burch (Ch.), Federico (M.), Bertoldi (N.), Cowan (B.), Shen (W.), Moran (Ch.), Zens (R.), Dyer (Ch.), Bojar (O.), Constantin (A.) et Herbst (E.), 2007 : « MOSES: Open source toolkit for statistical machine translation », dans *Proceedings of the ACL 2007 Demo and Poster Sessions*, Prague, June 2007, p. 177-180.
- Marcu (D.) et Munteanu (D. S.), 2005 : « Statistical Machine Translation: An English-Romanian Experiment », *EUROLAN 2005*.
- Och (F. J.) et Ney (H.), 2003 : « A Systematic Comparison of Various Statistical Alignment Models », dans *Computational Linguistics*, vol. 29, n° 1, March 2003, p. 19-51.
- Steinberger (R.), Pouliquen (B.), Widiger (A.), Ignat (C.), Erjavec (T.), Tufiş (D.) et Varga (D.), 2006 : « The JRC-Acquis: A Multilingual Aligned Parallel Corpus with 20 + Languages », dans *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, May 2006, p. 2142-2147.
- Todiraşcu (A.), Heid (U.), Ştefănescu (D.), Tufiş (D.), Gledhill (C.), Weller (M.) et Rousselot (F.), 2008 : « Vers un dictionnaire de collocations multilingue », dans *Cahier de Linguistique*, vol. 33, n° 1, Louvain, août 2008, p. 161-186.
- Tufiş (D.), 2002 : « A cheap and fast way to build useful translation lexicons », dans *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002)*, Taipei, August 2002, p. 1030-1036.
- Tufiş (D.), Ion (R.), Ceaușu (A.) et Ştefănescu (D.), 2006 : « Improved Lexical Alignment by Combining Multiple Reified Alignments », dans Ishida (T.), Fussell (S. R.) et Vossen (P.T.J.M.), éd., *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL2006)*, April 2006, Trento, Association for Computational Linguistics, p. 153-160.
- Tufiş (D.) et Koeva (S.), 2007 : « Ontology-supported Text Classification based on Cross-lingual Word Sense Disambiguation », dans Masulli (F.), Mitra (S.) et Pasi (G.), éd., *Applications of Fuzzy Sets Theory. 7th International Workshop on Fuzzy Logic and Applications (WILF 2007)*, volume 4578 of *Lecture Notes in Artificial Intelligence*, September 2007, Springer-Verlag, p. 447-455.

Tufiş (D.), Koeva (S.), Erjavec (T.), Gavrilidou (M.) et Krstev (C.), 2008 : « Building Language Resources and Translation Models for Machine Translation focused on South Slavic and Balkan Languages », dans Tadić(M.), Dimitrova-Vulchanova (M.) et Koeva (S.), éd., *Proceedings of the Sixth International Conference Formal Approaches to South Slavic and Balkan Languages (FASSBL 2008)*, Dubrovnik, September 25-28, p. 145-152.

Serge Verlinde,
Institut des langues vivantes,
K.U.Leuven,
Leuven,
Belgique,
serge.verlinde@ilt.kuleuven.be.

La *Base lexicale du français* : un outil lexicographique multi-usage

Résumé

La *Base lexicale du français* (BLF) est un outil lexicographique destiné à l'apprentissage du français qui répond aux exigences que l'on peut imposer à un ouvrage de référence développé pour le Web. Il regroupe de multiples fonctionnalités qui mènent l'utilisateur à une information précise. Pour réaliser cela, un tout nouveau type d'interface a été conçu à partir des besoins des utilisateurs. La *BLF* propose en outre un système d'aide à la lecture, à la traduction et à la rédaction contextualisée.

Mots-clés : lexicographie ; dictionnaire ; Web ; apprentissage ; français.

Introduction

Dans un livre récent, Tarp (2008 : 12) pose qu'un ouvrage de référence lexicographique basé sur les technologies de l'information modernes devrait prendre la forme d'un moteur de recherche avec accès à une base de données ou au Web. Cet outil, qu'il nomme *leximat*, devrait permettre à tout utilisateur ayant un besoin communicatif ou cognitif spécifique d'accéder par une recherche active ou passive aux données lexicographiques, à partir desquelles il pourrait extraire l'information dont il a besoin dans un contexte particulier.

Les dictionnaires informatisés, basés dans la plupart des cas sur des versions papier antérieures, ont amélioré l'accès à leur contenu à l'aide de fonctionnalités de recherche particulièrement puissantes (recherches de type flou basées sur la prononciation ou admettant un degré d'erreur dans la requête, recherches plein texte) ou en combinant diverses ressources, comme sur le site du *CNRTL* (Centre national de ressources textuelles et lexicales). Toutefois, malgré ces adaptations, ces ressources n'arrivent pas toujours à répondre rapidement aux questions fréquentes que se posent par exemple les apprenants, telle celle de savoir quelle est l'orthographe exacte des mots à flexion irrégulière ou des formes verbales irrégulières (Pruvoſt 2003). La consultation de ces dictionnaires informatisés, qui diffère somme toute peu de la consultation des dictionnaires papier, requiert en outre des aptitudes spécifiques, dont on sait qu'elles manquent souvent chez les utilisateurs (Miller 2006).

Notre contribution a pour objectif d'illustrer comment concevoir un outil lexicographique qui répondrait à ces nouvelles exigences en permettant à l'utilisateur d'accéder de façon plus rapide et efficace à l'information requise, tout en partant de ses besoins. La nouvelle interface de la *Base lexicale du français* (BLF), un dictionnaire d'apprentissage en ligne pour le français, accessible librement et destiné à des apprenants de niveau intermédiaire ou avancé, servira d'exemple. La BLF tire ses informations de notre propre description lexicographique des 6 500 lemmes les plus fréquents du français, le *Dictionnaire d'apprentissage du français langue étrangère ou seconde* (DAFLES), complétée par des informations de nombreux sites auxquels nous renvoyons à l'aide de raccourcis. Les sites intégrés de cette façon dans la BLF sont inventoriés sur la page d'accueil du site.

1 Des besoins communicatifs et cognitifs à l'architecture de l'interface

Diverses sources (Hausmann (1977 : 144) déjà et Tarp (2008 : 77) très récemment) ont proposé des listes de questions que peuvent se poser les utilisateurs de dictionnaires. En développant ces idées, nous croyons pouvoir distinguer six grands types de consultations potentiels pour toute personne confrontée à une question relative à un mot, à un groupe de mots (unités polylexicales: collocations, locutions; expressions, proverbes) ou au lexique en général:

- recherche d'informations: l'utilisateur part d'un mot ou d'une unité polylexicale et veut des informations en rapport avec ce mot ou cette unité polylexicale (genre, orthographe, sens, traduction, etc.);
- recherche d'une traduction vers une langue étrangère: pour des mots, mais également pour des unités polylexicales;
- vérification: l'utilisateur veut vérifier si quelque chose se dit, si une traduction qu'il a en tête est correcte;
- aide à la lecture, à la traduction et à la rédaction contextualisée;
- apprentissage/enseignement: l'utilisateur désire savoir comment le lexique est structuré, comment il peut éviter des erreurs courantes;
- apprentissage/exercices: l'utilisateur veut s'entraîner en faisant des exercices.

Il s'agit essentiellement de besoins communicatifs dans les deux premiers cas et de besoins cognitifs dans les deux derniers. La fonctionnalité de vérification et l'aide contextualisée peuvent répondre aux deux types de besoins.

L'avantage d'un système qui fusionne toutes ces fonctionnalités en une application, comparé à la dispersion actuelle des ressources et à la multiplication des sites, réside dans un gain de temps évident pour l'utilisateur et dans une information qui a fait l'objet d'une sélection. En outre, le fait que l'informatique permet de traiter des textes entiers signifie que l'information colle parfaitement au texte et que la consultation ne doit plus se faire mot par mot.

Le type de consultation détermine l'accès à la *BLF*, comme le montre la page d'accueil (figure 1), où sont visualisées les six options présentées ci-dessus.

HOME WIE IS WIE ORGANIGRAM CONTACT INTRANET ENGLISH

KATHOLIEKE UNIVERSITEIT
LEUVEN

Over K.U.Leuven | Onderwijs | Onderzoek | Toekomstige studenten | Studenten | Alumni | Faciliteiten | Bibliotheken | Faculteiten & Departementen | Internationaal

Leuven Language Institute

Lexical Database for French (Base lexicale du français - BLF) - new site

(Almost) everything you always wanted to know about... French words

Get information on

- a () **word**
Meaning, gender, use of prepositions, translation, ...
- a () **word combination/expression**
... *for lors de, une ambiance règne, à tout prix, ...*

Get the translation of

- a () **word to ()**
- a () **word combination/expression to ()**
een vraag stellen, ask a question, hacer una pregunta > ? une question

Verify

- the **use** of a (sequence of) word(s)
what do you say: *apparaître (à or sur) l'écran? espérer (de or -) faire ...* → check on the web
- if a **translation** is correct
salarié > salaried person?

Help me

- to **understand** a (short) text
"Beta" version: it's still a work in progress
- to **translate** a (short) text
Coming very soon...
- to **write correctly** a (short) text
Coming soon...

Learn

- how to **express an idea**
Something starts, something is intense, a large amount of something, words and word combinations about economics, ...
- how to **combine words** (correctly) to form sentences
salary > to earn a - | employer, company, colleagues, ...
- how to **avoid common errors**
Use of prepositions, position of the adjective, gender, ...

Do (a lot of) exercises

- on **all aspects of the vocabulary** of ()
On verb forms, gender, use of prepositions, word combinations, ...

Interface

- In English
Coming soon
- en français
- In het Nederlands

Help us to improve this tool

- Did you find the information you needed?
 yes no
- Why not?
- Suggestions?
-

Accented letters and special characters

- à é ò ó
- â ê ö ü
- ç è é ö ü
- ù ü ö ü
- ø
- œ
- Å
- &
- &

Figure 1

À chaque accès sont liées des possibilités de requête différentes. La figure 2 visualise la liste de questions que peut soumettre l'utilisateur à la base de données en sélectionnant l'option Informations/mot simple.

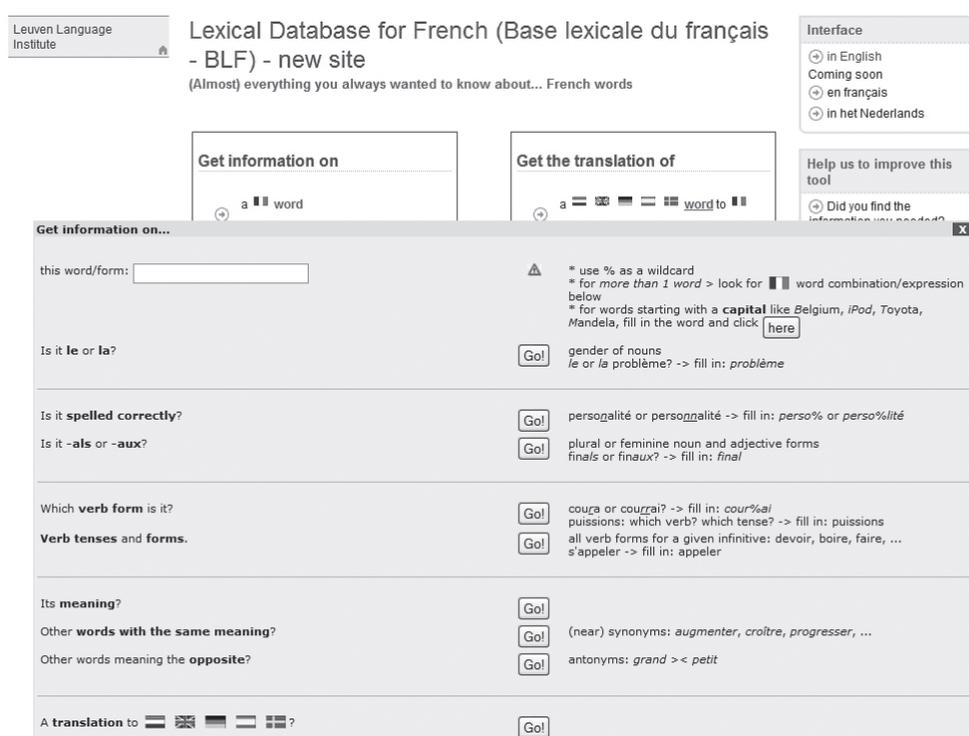


Figure 2

Les menus proposent à l'utilisateur de

- compléter une zone de saisie en tapant un mot ou une unité polylexicale et en cliquant sur le bouton renvoyant à la question qu'il se pose. Le dictionnaire lui retourne alors l'information précise qui répond à sa question: le genre du mot tapé, son sens, sa traduction, etc. ;
- cliquer directement sur un bouton qui permet d'extraire des informations structurées de la base de données: liste d'adjectifs qui expriment une idée d'intensification, liste des locutions prépositives, etc. ;
- coller un texte dans une zone de saisie pour obtenir une aide à la lecture, à la traduction ou à la rédaction contextualisée.

Le type d'informations reçues en réponse à la requête soumise peut prendre des formes très variées (figures 3 et 4).

DAFLES (Dictionnaire d'apprentissage du français langue étrangère ou seconde)

Le dictionnaire est en cours de réalisation. Merci de signaler les erreurs rencontrées. Pour d'autres applications du Dafles, cliquez ici.

●

pour .: problème :., voir l'un des mots suivants:

problème
nom

♂ masculin singulier un - le - au - du problème

♂♂ masculin pluriel les - aux - des problèmes

Figure 3

La figure 4 illustre le type d'informations contextualisées que l'utilisateur obtient en soumettant un texte entier à la fonctionnalité d'aide à la lecture, sur laquelle nous revenons à la section 5.

Base lexicale du français
Interfacultair Instituut voor
Levende Talen

ReadAID - aide à la lecture

127 mots

Reprise des festivités du Nouvel_an, soulagement des touristes, et réouverture des grands magasins, Bang... out doucement à la normale, durant l' après-midi du 14 avril, après une nuit de troubles qui m... ment de dégénérer .

Encer... reprendre
-> sens

Gover... -> synonymes

quarti... es "chemises rouges ", les manifestants antigouvernementaux retranchés à

journa... ège du gouvernement, ont mis un terme, en_début_d' après midi, à l' occupation du

... t par des barricades depuis trois semaines. Accompagnés par leurs partisans et des

... du mouvement se sont rendus eux-mêmes au siège de la police métropolitaine .

Analy... e texte réalisés en 1.61 secondes

-> prononciation



Figure 4

L'architecture de la *BLF* repose donc sur une modification fondamentale de l'accès classique à la description lexicale. Au lieu de proposer, comme dans les dictionnaires informatisés (et papier également), un accès unique (zone de saisie ou liste alphabétique d'entrées) pour obtenir un pavé de texte à décoder (l'article du dictionnaire), la *BLF* offre un nouveau type d'interface orientée utilisateur avec divers chemins d'accès qui mènent à une information bien déterminée. Si l'utilisateur désire toutefois un complément d'information, il peut, grâce aux liens sur les mots, accéder à partir de n'importe quel écran à un article de dictionnaire classique.

2 De l'interface à une description lexicographique multivariée

Gouws (2007 : 66) souligne à juste titre que le lexicographe a tendance à surestimer les capacités linguistiques des utilisateurs de dictionnaires. Même s'il faut se rendre à l'évidence que la langue est un phénomène complexe et que, par définition, sa description le sera aussi, il faut bien admettre, par exemple, que l'utilisation de métalangage ou d'abréviations multiples rend la description lexicographique parfois peu transparente. C'est pour cette raison que nous optons, lorsque cela est possible, pour une présentation plus « didactisée », par exemple en mentionnant les articles auprès des substantifs pour faciliter la mémorisation du genre (figure 3).

On retrouve sous *apprentissage/enseignement* toute une section consacrée à une série de particularités du français qui posent problème aux allophones : place de l'adjectif, patrons de constructions syntaxiques pour les noms, adjectifs et verbes, etc. Sont ainsi rassemblées en un lieu des informations éparpillées dans les dictionnaires ou traitées souvent trop sommairement dans les grammaires. Pour la place de l'adjectif, par exemple, l'utilisateur trouvera sur une page l'ensemble des adjectifs dont la position par rapport au nom donne lieu à un sens différent. La description pour chaque adjectif se fait en suite sous la forme de deux colonnes, l'une regroupant les sens où l'adjectif est préposé au nom, l'autre où l'adjectif est postposé.

La description lexicographique de la *BLF* se différencie en outre du contenu que l'on trouve dans les dictionnaires pour apprenants traditionnels (langue seconde ou langue étrangère : Rey-Debove 1999 par exemple) par une attention toute particulière portée à l'aide à l'encodage. Cette aide se traduit essentiellement par toute une série de données sur la combinatoire des mots (Verlinde, Binon et Selva 2006), mais aussi par un inventaire des réalisations lexicales prototypiques des actants pour chaque sens des verbes (Verlinde *et al.* 2004). Cet inventaire se présente sous la forme de tableaux, les schémas actanciels. Pour chaque mot prototypique mentionné dans ces tableaux, l'utilisateur peut obtenir une série de synonymes ou d'hyponymes et découvrir toute une série de réalisations lexicales possibles dans cette position.

3 Exploitation des ressources du Web

La nomenclature de la *BLF* se limite aux 6 500 lemmes les plus fréquents tirés d'un corpus de textes journalistiques (Verlinde et Selva 2001). La description lexicographique de cet ensemble de mots n'est pas encore tout à fait complétée. Heureusement, ces dernières années, on a vu apparaître sur le Web de nombreuses ressources électroniques pour le français (Habert 2005). Dans la *BLE*, nous avons tenté d'exploiter au maximum ces ressources pour compléter les lacunes de notre propre description. Les informations disponibles sur ces sites sont immédiatement accessibles par l'intermédiaire de raccourcis qui ouvrent ces sites externes à la page voulue (figure 5 pour la famille de mots *PROBLÈME* sur le site *Orthonet* à partir de la requête *Mots avec le même sens* de la figure 2).



Résultat de votre recherche

problème n.masc.

(question à résoudre)
 * un problème de géométrie et sa solution
 * avoir des problèmes
 (des difficultés)
 * le problème social, politique, moral..
 * soulever, poser un problème délicat
 * résoudre un problème
 (on évitera: "solutionner")

problématique adj. ETn.fém.

* un projet problématique
 (de réussite douteuse)
 * la problématique (ensemble des problèmes)
 du progrès technique

Figure 5

Le recours à des ressources externes permet également de compléter la *BLF* pour les traductions. Outre les sites de dictionnaires bi- ou multilingues, on trouve également aujourd'hui des corpus parallèles en accès libre (projet *Opus*, Tiedemann et Nygaard 2004), qui se révèlent particulièrement riches du point de vue des variantes de traduction que l'on peut y découvrir. Les corpus disponibles sur le site d'*Opus* sont alignés au niveau de la phrase et du mot. Les variantes de traduction peuvent être retrouvées dans ces deux types de ressources.

Grâce à l'intégration d'une multiplicité d'informations et de ressources, la *BLF* est un outil qui s'adapte facilement au niveau des utilisateurs. Tant un apprenant de niveau intermédiaire, qui a encore des doutes quant à la morphologie ou au genre des mots, qu'un linguiste, qui veut se livrer à une analyse détaillée de la synonymie à partir de la combinatoire des mots, y trouveront des informations pertinentes.

4 Vers une aide en ligne intégrée

C'est toutefois avec la fonctionnalité d'aide à la lecture, à la traduction et à la rédaction que la *BLF* se distingue le plus d'autres ressources lexicales en ligne (voir aussi la figure 4). L'aide à la lecture et l'aide à la traduction procèdent de la même façon : un texte soumis est analysé et renvoyé enrichi d'une série de liens vers des pages de la *BLF* (sens, traduction, synonymes, etc.) pour chaque mot ou unité polylexicale. Ces informations apparaissent dans des écrans qui surgissent au moment où l'utilisateur glisse la souris sur un mot du texte.

L'aide à la rédaction, qui est en cours d'élaboration, a pour but essentiel de favoriser chez l'apprenant une prise de conscience des erreurs récurrentes à l'écrit et de développer un moyen de les éviter en systématisant le contrôle de celles-ci. Dans un premier temps, l'application identifie dans un texte soumis les occurrences de patrons syntaxiques et lexicaux spécifiques très variés donnant souvent lieu à des erreurs chez les apprenants (place de l'adjectif, emploi des prépositions, choix entre l'imparfait et le passé composé, etc.). Ensuite, elle regroupe et visualise toutes les occurrences de ces patrons. L'apprenant peut ensuite contrôler ces occurrences à l'aide d'une sorte de mini-grammaire, enrichie dans la mesure du possible de données que nous tirerons de la *BLF* et qui apparaîtront sur le même écran (Verlinde *et al.* à paraître). Sur le plan lexical, l'application suggérera à l'utilisateur des synonymes et des combinaisons de mots variés pour les mots relevés dans le texte. Ces informations peuvent l'aider à varier et à enrichir son texte.

Contrairement au correcteur grammatical de Word, par exemple, l'application ne procède donc pas à une correction immédiate du texte, mais veut stimuler la capacité d'autocorrection et d'autoformation de l'apprenant.

5 Développements futurs

Un dispositif de traces enregistre tout le trafic sur le site par l'identification de plus de 250 chemins que peut suivre l'utilisateur à l'intérieur de la base et vers les sites externes. Par une analyse de ces traces, nous espérons pouvoir cibler davantage les habitudes de consultation des utilisateurs afin d'optimiser l'outil, tant en ce qui concerne son contenu qu'en ce qui concerne les voies d'accès aux informations (Verlinde et Binon, à paraître).

Nous voudrions également tirer profit de la présence sur le Web de nombreuses ressources pour d'autres langues que le français et les intégrer dans notre interface, qui deviendrait ainsi une sorte de site portail lexical à accès unique, standardisé pour tout problème lié au lexique.

Conclusion

La lexicographie a définitivement laissé derrière elle *l'artisanat des dictionnaires papier* (Rey 2008) pour entrer dans l'ère du Web. Les opportunités qu'offre l'informatique, couplées à des descriptions lexicographiques enrichies, ouvrent de toutes nouvelles perspectives en matière d'intégration de ressources et de développement d'aides performantes. C'est tout au profit de l'utilisateur du « dictionnaire ».

Bibliographie

BLF, 2009 : *Base lexicale du français*, K.U.Leuven, Institut des langues vivantes, site consulté le 10 novembre 2009, ilt.kuleuven.be/BLF.

- CNRTL, 2009 : *Centre national de ressources textuelles et lexicales*, CNRS, ATILF, site consulté le 10 novembre 2009, www.cnrtl.fr.
- Gouws (R.), 2007 : « Sublemmata or main lemmata : a critical look at the presentation of some macro-structural elements », dans Gottlieb (H.) et Mogensen (J. E.), éd., *Dictionary visions, research and practice. Selected papers from the 12th international symposium on lexicography, Copenhagen 2004*, Amsterdam/Philadelphie : Benjamins, p. 55-70.
- Habert (B.), 2005 : *Instruments et ressources électroniques pour le français*. Gap/Paris : Ophrys, site consulté le 10 novembre 2009, www.limsi.fr/Individu/habert/Publications/InstrumentsEtRessourcesElectroniquesPourLeFrancais.html.
- Hausmann (F.-J.), 1977 : *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Tübingen : Niemeyer.
- Miller (J.), 2006 : « An investigation into the effect of English learners' dictionaries on international students' acquisition of the English article system », dans *International Education Journal*, vol. 7, n° 4, p. 435-445, site consulté le 10 novembre 2009, ehlt.flinders.edu.au/education/iej/articles/v7n4/Miller/paper.pdf.
- Opus, 2009 : *Opus – an open source parallel corpus*, site consulté le 10 novembre 2009, urd.let.rug.nl/tiedeman/OPUS/.
- Orthonet, 2009, CILE, site consulté le 10 novembre 2009, orthonet.sdv.fr/.
- Pruvošt (J.), 2003 : « *Some lexicographic concepts stemming from a French training in lexicology* », site consulté le 10 novembre 2009, kictionaries.com/newsletter/kdn11-03.html.
- Rey (A.), 2008 : *De l'artisanat des dictionnaires à une science du mot. Images et modèles*, Paris : Colin.
- Rey-Debove (J.), 1999 : *Dictionnaire du français*, Paris : CLE International/Le Robert.
- Tarp (S.), 2008 : *Lexicography in the borderland between knowledge and non-knowledge*, Tübingen : Niemeyer.
- Tiedemann (J.) et Nygaard (L.), 2004 : « The OPUS corpus », dans *Proceedings of the fourth international conference on language resources and evaluation (LREC'04). Lisbon, Portugal, May 26-28*, site consulté le 10 novembre 2009, stp.ling.uu.se/~joerg/paper/opus_lreco4.pdf.
- Verlinde (S.), Binon (J.) et Selva (T.), 2006 : « Corpus, collocations et dictionnaires d'apprentissage », dans *Langue française*, vol. 150, p. 85-98.
- Verlinde (S.) et Binon (J.), à paraître : « Monitoring dictionary use in the electronic age », Actes du 14^e colloque Euralex (Leeuwarden, 2010).
- Verlinde (S.), Paulussen (H.), Slootmaekers (A.) et De Wachter (L.), à paraître : « La conception d'un didacticiel d'aide à la rédaction », dans *Revue française de linguistique appliquée*.
- Verlinde (S.) et Selva (T.), 2001 : « Nomenclature de dictionnaire et analyse de corpus », dans *Cahiers de lexicologie*, vol. 79, p. 113-139.
- Verlinde (S.), Selva (T.), Binon (J.) et Petit (G.), 2004 : « Les schémas actanciels dans le dictionnaire : point de convergence entre la morphologie et la sémantique lexicale », dans Williams (G.) et Vessier (S.), éd., *Proceedings of the eleventh EURALEX international congress*, Lorient, vol. II, p. 427-437.

Patrice Pognan,
LALIC (U. Paris-Sorbonne et INALCO),
INALCO,
Paris,
France,
mcertal@wanadoo.fr.

Bases de données pour l'étude d'un groupe de langues : l'exemple des langues slaves de l'Ouest

Résumé

Nous développons un prototype multifonctionnel et multilingue de bases de données d'après lequel pourront être dupliquées les bases de données de différentes langues, actuellement berbère, albanais, slovaque, slovène et tchèque. Parmi les nombreuses spécificités de nos bases de données, nous nous attardons à la description de l'appareil multilingue et comparativiste qui intègre une dimension importante consacrée à la phonologie diachronique, ce qui ouvre de nombreuses applications en synchronie (didactique et traitement automatique).

Mots-clés : bases de données ; système linguistique des langues slaves ; multilinguisme ; grammaires synchronique et diachronique ; applications pédagogiques.

Introduction

Nous assistons depuis quelques années à la construction systématique d'outils de conservation, d'organisation, de consultation et d'exploitation des données, en particulier avec le renouveau des corpus et des bases de données. Au-delà de la consultation de corpus en ligne et de la constitution de nos corpus à des fins de développement et de tests de programmes de traitement automatique des langues, nous cherchons à développer de véritables « chantiers de données » qui puissent répondre simultanément à plusieurs préoccupations :

- apporter au traitement automatique des langues le maximum d'exhaustivité et de finesse dans les descriptions linguistiques ;
- fournir un support à l'apprentissage des langues dans un cadre multilingue, fournir en même temps la culture grammaticale nécessaire à un accès à une intercompréhension exogène ;
- être à l'origine de la production de dictionnaires, de grammaires et de méthodes, en particulier d'auto-apprentissage.

Nous nous sommes rapidement orienté vers les bases de données pour la pérennité de données engrangées en un lieu unique, pour la souplesse de l'outil et pour la structuration de la vision linguistique qu'elles proposent. Par souci d'efficacité, nous développons un outil multifonctionnel et multilingue, sorte de matrice générale que l'on peut ensuite dupliquer pour l'étude lexicale et grammaticale de langues particulières au prix d'aménagements minimes. Cet outil doit permettre de rassembler des connaissances de plusieurs types :

- des connaissances lexicales pour la réalisation de dictionnaires ;
- des connaissances grammaticales synchroniques et diachroniques situées à différents niveaux (phonétique, phonologie, morphologie, syntaxe, sémantique) ;
- des connaissances spécialisées (scientifiques, médicales, techniques).

Ces bases de données ont le français comme pivot (bases de données langue étrangère-français, puis français-langue étrangère), mais aussi comme langue d'interface personne-machine. Le prototype élaboré est taillé pour pouvoir rendre compte de langues indo-européennes, de langues chamito-sémitiques et de langues agglutinantes, notamment ouralo-altaïques.

Les modèles dont le développement est le plus avancé concernent le berbère (dans ses variétés marocaines tachelhit et tamazight) pour les langues chamito-sémitiques ; l'albanais, le slovaque, le slovène et le tchèque pour les langues indo-européennes. Le traitement de langues agglutinantes est actuellement préparé par l'étude des structures du turc et de l'azéri. Tous ces modèles ont la même structure plus ou moins développée suivant les langues. Ainsi, un seul onglet couvre la morphologie du berbère (figure 1).

Nous consacrerons cet article aux aspects multilingues et ne reviendrons pas sur la structuration de ces bases de données qui peut être consultée dans Pognan (2009a et 2009b).

Nous avons choisi de présenter les langues slaves qui sont un bon exemple de l'unité d'un système linguistique avec une dimension diachronique tangible que ne possèdent pas, p. ex., les langues berbères.

43

LEXIQUE DE BERBÈRE TACHELHIT

Lexie tamgart ⵜⴰⵎⴳⴰⵔⵜ

GÉNÉRALITÉS
 LEXIQUE
 MORPHOLOGIE
 AUTRES LANGUES
 SOURCES

formes verbales simples

impératif MGR

FV1 aoriste

FV2 inaccompli aoriste intensif

FV3 accompli

FV4 accompli négatif

prétérit

prétérit négatif

formes verbales dérivées

pos. 3 pos. 2 pos. 1

factitif p3_S p2_S p1_S

réciproque p3_M p2_M p1_M

réfléchi, expressif p3_N p2_N p1_N

passif p3_T p2_T p1_T

formes verbo-nominales et nominales

genre féminin valeur particulière

43 tamgart

| SINGULIER | | PLURIEL | |
|----------------------|----------------------|----------------------|----------------------|
| masculin état libre | <input type="text"/> | masculin état libre | <input type="text"/> |
| masculin état annexé | <input type="text"/> | masculin état annexé | <input type="text"/> |
| fémnin état libre | <input type="text"/> | fémnin état libre | <input type="text"/> |
| tamgart | <input type="text"/> | tamgarin / tamgarin | <input type="text"/> |
| fémnin état annexé | <input type="text"/> | fémnin état annexé | <input type="text"/> |
| tmgart | <input type="text"/> | tmgarin | <input type="text"/> |

Enr : 1 sur 1

autre genre amgar / argaz

Figure 1

1 Le système linguistique commun aux langues slaves

Les langues slaves sont un exemple typique de langues à flexion externe. Ces flexions sont marquées dans la sphère nominale par sept cas (le nominatif, le génitif, le datif, l'accusatif, le vocatif, le locatif et l'instrumental – l'usage de ce système est variable : totale ou presque pour le bulgare et le macédonien, elle est généralement circonscrite au vocatif, le cas de l'appel, le maillon le plus faible). Le système slave possède trois nombres (singulier, duel et pluriel). Comme dans d'autres langues, c'est le duel qui subit les outrages du temps, allant de quelques traces ténues à la plénitude du système en slovène et dans les deux sorabes. Les questions de genre sont plus originales. On possède en effet trois genres, le masculin, le féminin et le neutre, mais il existe des particularités au niveau du masculin. Il y a eu de très nombreux débats sur la question de savoir si certaines langues slaves possèdent trois, quatre ou cinq genres. Notre présentation répond d'elle-même à la position que nous avons adoptée : trois genres avec la présence de trois sous-genres masculins, le masculin inanimé, le masculin animal et le masculin humain. Il semble que le nombre de sous-genres soit croissant avec l'évolution des langues slaves. Le polonais, le haut- et le bas-sorabe possèdent les trois sous-genres.

Au niveau du verbe, dont la caractéristique la plus patente est la présence de l'aspect, les temps simples du passé (imparfait et aoriste) ont souvent laissé la place aux seuls temps composés. Les sorabes conservent les temps simples du passé.

Une caractéristique essentielle des langues slaves est l'autodétermination de la morphologie par la forme, ce que J.-M. Zemb dans ses cours du Collège de France appelait « l'auto-éclairage des mots ». Les paradigmes de la flexion des substantifs sont, en gros, donnés par l'orientation suivante : un substantif terminé en consonne est un masculin, un substantif terminé en « -a » est un féminin, et en « -o » un neutre. Bien entendu, il existe des croisements qui masquent quelque peu cet ordonnancement : il existe des féminins consonantiques et des masculins vocaliques. Mais dans cet ordre d'idée, un adjectif dit « dur » en tchèque est un mot qui se termine en « -ý » au masculin nominatif singulier et éventuellement à l'accusatif. Dans la langue tchèque, tout mot se terminant en « -ý » est un adjectif en dehors de trois exceptions. On comprendra alors aisément pourquoi nous avons choisi des traitements par reconnaissance automatique de formes pour l'analyse du tchèque (Pognan 1998 et Pognan 2007).

2 Le système linguistique général, cadre de description des langues individuelles

Nous enseignons la grammaire du tchèque par rapport à la dimension du système slave. Des éléments de comparaison avec les autres langues (une ou plusieurs) du même système permettent de se familiariser facilement avec ces autres langues, mais aussi d'intérioriser le système linguistique slave pour une meilleure compréhension des phénomènes.

3 La base de données, outil pédagogique

Les bases de données possèdent de très nombreuses applications, mais il n'était vraisemblablement pas prévu d'utiliser l'interface homme-machine à des fins pédagogiques. C'est pourtant ce qui nous a décidé à entreprendre ce projet essentiellement avec des bases du logiciel *Access*, retenu pour la qualité de ses formulaires relativement faciles à mettre en œuvre.

La définition de la base de données est celle du système général et non pas celle des langues individuelles. Celles-ci sont un sous-ensemble plus ou moins vaste du système linguistique pouvant aller jusqu'aux limites mêmes de ce système.

Nous donnerons de nombreux extraits des bases de données par captures d'écran parce que nous pensons que l'aspect visuel est un élément pédagogique majeur qui participe largement à l'apprentissage.

3.1 La flexion nominale

D'après la présentation du système linguistique slave, il est facile de conclure que la flexion du substantif slave est pour chacun des genres une matrice de 3 nombres x 7 cas, soit un ensemble de 21 éléments. Nous donnons ci-dessous la flexion du substantif slovène féminin « *gostilna* » auberge (figure 2) :

9

LEXIQUE SLOVÈNE

Lexie **gostilna**
gostilna

GÉNÉRALITÉS | **LEXIQUE** | DÉRIVATION VERBALE | PARADIGMES SUBSTANTIVAUX | DÉRIVATION NON VERBALE | FLE

classe lexicale substantif | classification 06 - miza | genre féminin | valeur particulière

comparatif | comparatif composé | superlatif | superlatif composé

3 gostilna

flexion substantivale

| singulier - ednina | | duel - dvojina | | pluriel - množina | |
|--------------------|----------|------------------|------------|-------------------|------------|
| flexion | | flexion | | flexion | |
| imenovalnik 1: N | gostilna | imenovalnik 1: N | gostilni | imenovalnik 1: N | gostilne |
| rodilnik 2: G | gostilne | rodilnik 2: G | gostiln | rodilnik 2: G | gostiln |
| dajalnik 3: D | gostilni | dajalnik 3: D | gostilnama | dajalnik 3: D | gostilnam |
| tožilnik 4: A | gostilno | tožilnik 4: A | gostilni | tožilnik 4: A | gostilne |
| 5: V | | 5: V | | 5: V | |
| mestnik 6: L | gostilni | mestnik 6: L | gostilnah | mestnik 6: L | gostilnah |
| orodnik 7: I | gostilno | orodnik 7: I | gostilnama | orodnik 7: I | gostilnami |

Entr : I sur 1

Figure 2

Le slovène possède le duel de manière généralisée, c'est-à-dire dans la conjugaison des verbes, la déclinaison des substantifs, des adjectifs, des pronoms, ce qui vaut dans la base de données l'activation des trois nombres. Mais le slovène ne possède pas de vocatif: on voit apparaître l'absence du vocatif par le rang naturellement affecté à celui-ci (le 5^e) qui est vide, inactivé et inaccessible. Nous avons délibérément choisi de marquer visuellement cette absence dans le formulaire par un à-plat qui contraste avec les champs actifs mis en relief.

Le tchèque, au contraire, possède un vocatif généralisé qui ne commence à disparaître que lorsqu'il porte sur un nom de famille. Le duel, par contre, n'existe plus que sous forme résiduelle (*deux, deux cents* et une dizaine de mots représentant des organes externes fonctionnant en paire). C'est le cas, par exemple des yeux:

- [1] « oko » un œil, « oči » les yeux (mais deux), forme de duel conservée dans la langue actuelle. Cependant « oka », le pluriel existe pour des acceptions dérivées du sens premier: les lacets du braconnier, de petits lacs de montagne, les yeux de bouillon gras!

1

LEXIQUE TCHÈQUE

Lexie **zaručovat**

GÉNÉRALITÉS
 LEXIQUE
 DÉRIVATION VERBALE
 PARADIGMES SUBSTANTIVAUX
 DÉRIVATION NON VERBALE
 FLI

classe lexicale genre classification valeur particulière

comparatif comparatif composé superlatif superlatif composé

▶ (Auto) zaručovat

flexion substantivale

| | SINGULIER | | DUEL | | PLURIEL | |
|--------------|-----------|----------------------|---------|----------------------|---------|----------------------|
| | flexion | | flexion | | flexion | |
| nominatif | 1: | <input type="text"/> | 1: | <input type="text"/> | 1: | <input type="text"/> |
| génitif | 2: | <input type="text"/> | 2: | <input type="text"/> | 2: | <input type="text"/> |
| datif | 3: | <input type="text"/> | 3: | <input type="text"/> | 3: | <input type="text"/> |
| accusatif | 4: | <input type="text"/> | 4: | <input type="text"/> | 4: | <input type="text"/> |
| vocatif | 5: | <input type="text"/> | 5: | <input type="text"/> | 5: | <input type="text"/> |
| locatif | 6: | <input type="text"/> | 6: | <input type="text"/> | 6: | <input type="text"/> |
| instrumental | 7: | <input type="text"/> | 7: | <input type="text"/> | 7: | <input type="text"/> |

Entr : I sur 1

Figure 3

C'est la raison pour laquelle le formulaire présente le singulier et le pluriel comme pleinement valides et laisse le duel de la couleur du fond en principe non activé, mais accessible pour la notation des formes résiduelles (figure 3). Le slovaque va présenter exactement le même fonctionnement à ceci près qu'il ne possède plus de vocatif. Nous avons donc un croisement des deux tableaux précédents, pas de vocatif comme le slovène, duel résiduel comme le tchèque (figure 4):

LEXIQUE SLOVAQUE

Lexie **deň**

GÉNÉRALITÉS
 LEXIQUE
 DÉRIVATION VERBALE
 PARADIGMES SUBSTANTIVAUX
 DÉRIVATION NON VERBALE
 FLI

classe lexicale substantif
 genre m. inanimé
 classification 04 - stroj
 valeur particulière

comparatif
 comparatif composé
 superlatif
 superlatif composé

1 deň

flexion substantivale

| | SINGULIER | | DUEL | | PLURIEL | |
|--------------|-----------|----------------------|---------|----------------------|---------|----------------------|
| | flexion | | flexion | | flexion | |
| nominatif | 1: | <input type="text"/> | 1: | <input type="text"/> | 1: | <input type="text"/> |
| génitif | 2: | <input type="text"/> | 2: | <input type="text"/> | 2: | <input type="text"/> |
| datif | 3: | <input type="text"/> | 3: | <input type="text"/> | 3: | <input type="text"/> |
| accusatif | 4: | <input type="text"/> | 4: | <input type="text"/> | 4: | <input type="text"/> |
| vocatif | 5: | <input type="text"/> | 5: | <input type="text"/> | 5: | <input type="text"/> |
| locatif | 6: | <input type="text"/> | 6: | <input type="text"/> | 6: | <input type="text"/> |
| instrumental | 7: | <input type="text"/> | 7: | <input type="text"/> | 7: | <input type="text"/> |

Err : 14 1 sur 1

Figure 4

3.2 La conjugaison

Nous présenterons successivement la conjugaison du présent en tchèque (figure 5) et en slovène (figure 6), la construction en tchèque et en slovaque du lexème verbal du passé et le passé composé qui en découle.

LEXIQUE TCHÈQUE

Lexie **zaručovat**

GÉNÉRALITÉS
 LEXIQUE
 DÉRIVATION VERBALE
 PARADIGMES SUBSTANTIVAUX
 DÉRIVATION NON VERBALE
 FLI

généralités
 classification III-2 - kupovat
 infinitif zaručovat
 racine Ruč
 aspect ipf

série verticale
 aspect 1
 aspect 2

série transversale
 aspect 1 zaručovat
 ipf
 aspect 2 zaručit
 pf
 aspect 3

aspect 4
 aspect 5
 aspect 6

1 zaručovat

présent
 imparfait
 aoriste
 participe passé actif
 passé composé
 plus que parfait
 conditionnel
 impératif
 formes impersonnelles

présent

| singulier | | duel | | pluriel | |
|-----------|-----------|------|----------------------|---------|------------|
| 1: | zaručuji | 1: | <input type="text"/> | 1: | zaručujeme |
| 2: | zaručuješ | 2: | <input type="text"/> | 2: | zaručujete |
| 3: | zaručuje | 3: | <input type="text"/> | 3: | zaručují |

Err : 14 1 sur 1

Figure 5

La conjugaison du présent slovène, à la différence du présent tchèque, comporte un duel qu'il sera intéressant, pour la connaissance historique des langues slaves, de comparer à celui des langues sorabes.

LEXIQUE SLOVÈNE

Lexie **potrebovati**
potrebovati

GÉNÉRALITÉS | LEXIQUE | DÉRIVATION VERBALE | PARADIGMES SUBSTANTIVAUX | DÉRIVATION NON VERBALE | FLE

généralités classification kupovati infinitif potrebovati racine TReB aspect ipf

série verticale aspect 1 aspect 2

série transversale aspect 1 aspect 2 aspect 3

aspect 4 aspect 5 aspect 6

2 potrebovati

présent imparfait aoriste participe passé actif passé composé plus que parfait conditionnel impératif formes impersonnelles

présent

| singulier | duel | pluriel |
|----------------|----------------|----------------|
| 1: potrebujem | 1: potrebujeva | 1: potrebujemo |
| 2: potrebujesh | 2: potrebujeta | 2: potrebujete |
| 3: potrebuje | 3: potrebujeta | 3: potrebujete |

Figure 6

La série qui suit concerne la formation et l'accord en genre et nombre du lexème du passé (participe passé actif) à partir duquel sont formés les temps composés (passé composé, conditionnel - mais aussi de manière étonnante un futur en slovène). Il n'y a pas en tchèque et en slovaque d'expression du duel, ce qui est répercuté dans l'interface. Contrairement au polonais et aux sorabes, le tchèque et le slovaque n'ont pas distingué dans la conjugaison un masculin humain et un masculin animal (la déclinaison d'un nom d'animal masculin reste particulière en slovaque) : il n'y a qu'une opposition entre animé (un masculin morphologique qui désigne un masculin biologique animé - homme ou animal) et inanimé (masculin morphologique qui désigne une chose). Mais il existe une différence importante entre tchèque et slovaque : le tchèque continue à avoir au pluriel des formes distinguées suivant le genre alors que le slovaque (comme le russe) ne possède plus qu'une forme unique de pluriel, ce qui se transmet dans la formation des temps composés.

Lexème du passé (figures 7 et 8):

LEXIQUE TCHÈQUE

Lexie **zaručovat**

GÉNÉRALITÉS | **LEXIQUE** | DÉRIVATION VERBALE | PARADIGMES SUBSTANTIVAUX | DÉRIVATION NON VERBALE | FLI

généralités classification III-2 - kupovat | infinitif zaručovat | racine RuČ | aspect ipf
série verticale aspect 1 | aspect 2
série transversale aspect 1 zaručovat | ipf | aspect 2 zaručit | pf | aspect 3 | aspect 4 | aspect 5 | aspect 6

1 zaručovat

présent | imparfait | aoriste | participe passé actif | passé composé | plus que parfait | conditionnel | impératif | formes impersonnelles

participe passé actif

| singulier | | duel | | pluriel | |
|------------|------------|------------|--|------------|------------|
| m. animé | zaručoval | m. humain | | m. animé | zaručovali |
| m. animal | | m. animal | | m. animal | |
| m. inanimé | zaručoval | m. inanimé | | m. inanimé | zaručovaly |
| féminin | zaručovala | féminin | | féminin | zaručovaly |
| neutre | zaručovalo | neutre | | neutre | zaručovala |

Env : 14 | 1 | sur 1

Figure 7

LEXIQUE SLOVAQUE

Lexie **zaručovať**

GÉNÉRALITÉS | **LEXIQUE** | DÉRIVATION VERBALE | PARADIGMES SUBSTANTIVAUX | DÉRIVATION NON VERBALE | FLI

généralités classification 26 - pracovať | infinitif zaručovať | racine RuČ | aspect ipf
série verticale aspect 1 | aspect 2
série transversale aspect 1 zaručovať | ipf | aspect 2 zaručiť | pf | aspect 3 | aspect 4 | aspect 5 | aspect 6

1 zaručovať

présent | imparfait | aoriste | participe passé actif | passé composé | plus que parfait | conditionnel | impératif | formes impersonnelles

participe passé actif

| singulier | | duel | | pluriel | |
|------------|------------|------------|--|------------|------------|
| m. animé | zaručoval | m. humain | | m. animé | zaručovali |
| m. animal | | m. animal | | m. animal | |
| m. inanimé | zaručoval | m. inanimé | | m. inanimé | zaručovali |
| féminin | zaručovala | féminin | | féminin | zaručovali |
| neutre | zaručovalo | neutre | | neutre | zaručovali |

Env : 14 | 1 | sur 1

Figure 8

Passé composé (figures 9 et 10):

LEXIQUE TCHÈQUE

Lexie **zaručovat**

GÉNÉRALITÉS | LEXIQUE | DÉRIVATION VERBALE | PARADIGMES SUBSTANTIFS | DÉRIVATION NON VERBALE | FLI

généralités classification III-2 - kupovat infinitif zaručovat racine Ruč aspect ipf

série verticale aspect 1 aspect 2

série transversale aspect 1 zaručovat ipf aspect 2 zaručit pf aspect 3

aspect 4 aspect 5 aspect 6

1 zaručovat

présent imparfait aoriste participe passé actif passé composé plus que parfait conditionnel impératif formes impersonnelles

passé composé masculin animé masculin inanimé féminin neutre

| masculin animé | | masculin inanimé | | féminin | | neutre | |
|-----------------|-----------------|------------------|-------------------|---------|-----------------|--------|-----------------|
| masculin humain | | masculin animal | | | | | |
| 1 | zaručoval jsem | | zaručoval jsem | | zaručovala jsem | | zaručovalo jsem |
| 2 | zaručoval jsi | S | I zaručoval jsi | N | zaručovala jsi | G | zaručovalo jsi |
| 3 | zaručoval | | zaručoval | | zaručovala | | zaručovalo |
| 1 | | | | | | | |
| 2 | | D | U | E | | L | |
| 3 | | | | | | | |
| 1 | zaručovali jsme | | zaručovali jsme | | zaručovaly jsme | | zaručovala jsme |
| 2 | zaručovali jste | P | L zaručovali jste | U | zaručovaly jste | R | zaručovala jste |
| 3 | zaručovali | | zaručovali | | zaručovaly | | zaručovala |

Entr : 1 sur 1

Figure 9

LEXIQUE SLOVAQUE

Lexie **zaručovať**

GÉNÉRALITÉS | LEXIQUE | DÉRIVATION VERBALE | PARADIGMES SUBSTANTIFS | DÉRIVATION NON VERBALE | FLI

généralités classification 26 - pracovať infinitif zaručovať racine Ruč aspect ipf

série verticale aspect 1 aspect 2

série transversale aspect 1 zaručovať ipf aspect 2 zaručiť pf aspect 3

aspect 4 aspect 5 aspect 6

1 zaručovať

présent imparfait aoriste participe passé actif passé composé plus que parfait conditionnel impératif formes impersonnelles

passé composé masculin animé masculin inanimé féminin neutre

| masculin animé | | masculin inanimé | | féminin | | neutre | |
|-----------------|----------------|------------------|------------------|---------|----------------|--------|----------------|
| masculin humain | | masculin animal | | | | | |
| 1 | zaručoval som | | zaručoval som | | zaručovala som | | zaručovalo som |
| 2 | zaručoval si | S | I zaručoval si | N | zaručovala si | G | zaručovalo si |
| 3 | zaručoval | | zaručoval | | zaručovala | | zaručovalo |
| 1 | | | | | | | |
| 2 | | D | U | E | | L | |
| 3 | | | | | | | |
| 1 | zaručovali sme | | zaručovali sme | | zaručovali sme | | zaručovali sme |
| 2 | zaručovali ste | P | L zaručovali ste | U | zaručovali ste | R | zaručovali ste |
| 3 | zaručovali | | zaručovali | | zaručovali | | zaručovali |

Entr : 1 sur 1

Figure 10

3.3 La flexion automatique

Nous avons présenté les tables flexionnelles des substantifs et des verbes associées aux entrées. Ces informations permettent de produire des manuels de conjugaison

(de type Bescherelle) et de déclinaison. « Peupler » manuellement les champs correspondants serait un travail au-dessus des forces humaines. C'est pourquoi nous développons des composantes automatiques de génération flexionnelle. Nous donnons ci-dessous un aperçu d'une partie du programme de déclinaison des masculins inanimés en slovène (figure 11) :

```

Sg_1 [v] [Enter]
'modèle sir
If [Forms]![MOT]![modele de flexion] = "02 - sir" And IsNull([Sg-1]) Then
  For i = 0 To 3
    If mep(i) = Right$(lemme, Len(mep(i))) Then
      lemme2 = Left$(lemme, Len(lemme) - Len(mep(i))) + Right$(lemme, 1)
      Exit For
    Else
      lemme2 = lemme
    End If
  Next i
'singulier
[Sg-1] = lemme
[Sg-2] = lemme2 + des(0)
[Sg-3] = lemme2 + des(4)
[Sg-4] = [Sg-1]
[Sg-6] = [Sg-3]
[Sg-7] = lemme2 + "om"
'duel
[Du-1] = [Sg-2]
[Du-2] = lemme2 + "ov"
[Du-3] = lemme2 + "oma"
[Du-4] = [Du-1]
[Du-6] = lemme2 + "ih"
[Du-7] = [Du-3]
'pluriel
[Pl-1] = lemme2 + des(2)
[Pl-2] = [Du-2]

```

Figure 11

Parmi ces masculins, ceux terminés en « -ec », « -ek », « -en » et « -enj » perdent le « e » épenthétique, vocalisation d'un ancien *jer* vieux-slave, aux cas obliques (génitif, datif, locatif et instrumental). La morphologie des substantifs des langues slaves de l'Ouest demandera un traitement nettement plus complexe. Ces composantes, en cours d'élaboration, sont calculées en VBA, le langage de programmation interne à Microsoft Office. C'est le modèle slovène qui est le plus avancé pour la déclinaison et le modèle albanais¹ pour la conjugaison.

4 La base de données, outil de recherche philologique

L'expérience en traitement automatique des langues slaves de l'Ouest et de leur enseignement montre à quel point la prise en compte de la diachronie est importante. En analyse automatique, une solution prise en synchronie qui est l'image d'un phénomène historique est d'une fiabilité à toute épreuve. Pour l'enseignement de la langue contemporaine, l'éclairage de la diachronie apporte une compréhension

1 Objet d'un contrat AUF-LTT 2009-2010 « Bases de données pour l'étude grammaticale et lexicale des langues dans une visée multilingue » entre Paris, Fès et Tirana.

fortement améliorée des phénomènes linguistiques et emporte le total assentiment des apprenants (adultes) pour ce type de démarche.

Nous donnons ci-dessous une suite d'évolutions phonologiques subdivisée en trois périodes. Les changements les plus profonds et les plus généraux sont aussi les plus anciens. Signalons ici que la prise en compte de ces phénomènes de phonologie historique permet d'assurer une partie importante de la reconnaissance automatique des emprunts gréco-latins en tchèque (sans recours à un dictionnaire) (Pognan 1998 et Pognan 2007).

Voici cette suite (succincte) de phénomènes historiques :

- Du protoslave à la fin du x^e siècle :
 - Métathèse des liquides (évolution de : *tort, tolt, tert, telt*);
 - Contraction;
 - Disparition et vocalisation des *jers*;
 - Évolution des nasales (conservées uniquement en polonais).
- De la fin du x^e siècle à la fin du xiv^e siècle :
 - Passage de « g » à « h » pour le haut-sorabe, le tchèque et le slovaque;
 - Évolution du « r » mouillé en « ř » en tchèque;
 - Transformations de « a » en « ě » et de « u » en « i » derrière une consonne molle en tchèque;
 - Dépalatalisation en tchèque.
- De la fin du xiv^e siècle à la fin du xvi^e siècle (phénomènes concernant surtout le tchèque) :
 - Transformation du « u long » en diphtongue : de « ú » vers « ou »;
 - Contraction de « ie » en « í »;
 - Contraction de « uo » issu de « ó » vers « ů » en tchèque et « ô » en slovaque;
 - Changement de « aj » vers « ej » au sein d'une syllabe.

La division linguistique traditionnelle des langues slaves se fait en trois groupes : les langues slaves de l'Ouest (bas-sorabe et haut-sorabe sur le territoire de l'ex-Allemagne de l'Est, isolats slaves en milieu germanique du plus haut intérêt philologique, le polonais auquel il convient d'ajouter le cachoube et enfin le tchèque et le slovaque), les langues slaves de l'Est (russe, biélorusse et ukrainien auxquels il convient d'ajouter le ruthène) et les langues slaves du Sud que nous avons l'habitude de subdiviser en Est (bulgare et macédonien) et Ouest (ensemble serbe croate et slovène - probablement un hybride entre langues slaves du Sud et de l'Ouest).

Nous conservons scrupuleusement ces répartitions dans nos bases de données. Elles sont répercutées principalement par trois tables spécialisées. Ces tables, de forme matricielle, permettent de consigner pour chacune des langues slaves la trace des évolutions historiques et donc de consigner les éléments des écarts entre langues et de langue à langue. Nous donnons ci-après (figure 12) un extrait recomposé de ces tables :

| lexicographie comparative sur la base de l'évolution phonologique | | | | |
|---|--|--|--|--|
| 2 dan | slovène dan | tchèque den | slovaque deň | russe день |
| sens | | | | |
| contraction | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| métathèse | <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> |
| iers mou a | <input type="checkbox"/> mou <input type="checkbox"/> a | <input type="checkbox"/> e | <input type="checkbox"/> e | <input type="checkbox"/> mou <input type="checkbox"/> e |
| nasales | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| g/h | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| r mou | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| a/e u/i | <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> |
| dépallatalisation 1 | <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> |
| dépallatalisation 2 | <input checked="" type="checkbox"/> ñ <input type="checkbox"/> n | <input checked="" type="checkbox"/> ñ <input type="checkbox"/> n | <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> |
| ú/ou | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| ó | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| contraction ie | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| aj | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| mouillure x 2 e | <input type="checkbox"/> i <input type="checkbox"/> | <input type="checkbox"/> e <input type="checkbox"/> i <input type="checkbox"/> | <input type="checkbox"/> e <input type="checkbox"/> i <input type="checkbox"/> | <input type="checkbox"/> e <input type="checkbox"/> i <input type="checkbox"/> |

Figure 12

L'exemple donné ici concerne la vocalisation des *iers* vieux-slaves. En tchèque, par exemple, la vocalisation est en « e » que le *ier* originel ait été dur ou mou. En slovène, la situation est complexe pour ne pas dire inextricable. En effet, aussi bien l'ancien *ier* dur que l'ex-*ier* mou vocalisent tous les deux en « a » et en « e » : sur ce point, le passé du slovène est obscur. Il se trouve que le russe, le bulgare et le macédonien différencient dans la langue moderne l'héritage des *iers*. Ainsi, en russe un *ier* dur donne une vocalisation en « o » et un *ier* mou une vocalisation en « e ». Par la correspondance entre l'une de ces langues et le slovène, il sera possible d'établir une radiographie de la situation slovène qui, peut-être, pourrait déboucher sur des régularités. Dans le cas du mot slovène [2] « dan » jour, la vocalisation en « a » est issue d'un *ier* mou.

On remarquera dans cet exemple la dépallatalisation du tchèque et du slovène.

Mais ces phénomènes historiques ne sont pas les seuls que nous consignons dans ces tables. Nous tenterons de suivre à la trace un phénomène qui semble être passé quelque peu inaperçu. Il s'agit de ce que nous appelons « double mouillure ». Nous

avons pris conscience de ce phénomène par le rapprochement du tchèque et du haut-sorabe (ce qui met bien en lumière la validité des études comparatives). En tchèque, pour écrire un « e mouillé », nous avons deux solutions : soit « ě » c'est-à-dire un « e » surmonté d'un *háček*, par exemple [3] « *jed* » *poison*, soit une graphie « je » qui est obligatoire en début de mot. En haut-sorabe, le poison s'écrit « *jěd* ». En tchèque, le *háček* s'efface devant le « j », c'est le maillon faible (figure 13).

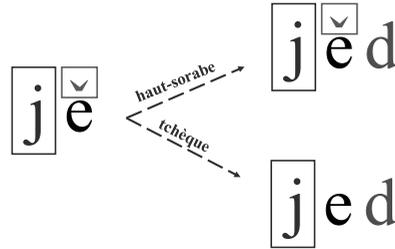


Figure 13

Ces tables permettent d'établir les régularités de transfert de langue à langue (figure 14) :

| régularités entre tchèque et slovaque | | | |
|--|-----------------|----------------|------------------------------|
| <i>phénomène phonologique concerné</i> | <i>slovaque</i> | <i>tchèque</i> | <i>traduction</i> |
| Evolution du « r » mouillé en « ř » | reč | řeč | « discours » |
| Transformations de « a » en « ě » et de « u » en « i » derrière une consonne molle | ovca pľúca | ovce plíce | « mouton » « poumons » |
| Dépalatalisation | hovorit' | hovořit | « parler » |
| Transformation de « ú » en « ou » | súd | soud | « tribunal » |
| Contraction de « ie » en « i » | miera | míra | « mesure » |
| Contraction de « uo » issu de « ó » vers « ů » en tchèque et « ô » en slovaque | pôvod | původ | « origine » |
| Changement de « aj » vers « ej » au sein d'une syllabe | naj- | nej- | <i>préfixe du superlatif</i> |

Figure 14

Il est très intéressant de constater que la forme tchèque est obtenue (en dehors de l'évolution du « o long ») par un pas d'évolution à partir d'une forme commune au tchèque et au slovaque.

La base de données, par l'intermédiaire des tables flexionnelles, donnera d'autres éléments de réflexion sur la grammaire historique. Montrons-le à l'aide de documents pédagogiques, tableaux de déclinaisons de substantifs à augment en tchèque [4] « *tele* », « le veau » (figure 15), et en haut-sorabe « *ćelo* » (figure 16).

| <i>tele</i> (<i>veau</i>) | singulier | duel | pluriel |
|-----------------------------|-----------|------|------------|
| nominatif | tel e | | tel at a |
| génitif | tel et e | | tel at |
| datif | tel et i | | tel at ům |
| accusatif | tel e | | tel at a |
| vocatif | tel e | | tel at a |
| locatif | tel et i | | tel at ech |
| instrumental | tel et em | | tel at y |

Figure 15

| <i>ćelo</i> (<i>veau</i>) | singulier | duel | pluriel |
|-----------------------------|-----------|-------------|------------|
| nominatif | ćel o | ćel ec i | ćel at a |
| génitif | ćel ec a | ćel ec ow | ćel at ow |
| datif | ćel ec u | ćel ec omaj | ćel at am |
| accusatif | ćel o | ćel ec i | ćel at a |
| vocatif | | | |
| locatif | ćel ec u | ćel ec omaj | ćel at ach |
| instrumental | ćel ec om | ćel ec omaj | ćel at ami |

Figure 16

Ces tableaux ont la forme que revêt la base de données. En tchèque, l'absence de duel ne permet pas de prendre conscience de la forme de l'augment du duel. Par le haut-sorabe, on apprend que c'est celui du singulier et qu'il apparaît à tous les cas. On apprend aussi que l'augment du pluriel est « -at- » de manière commune.

En haut-sorabe, la palatalisation de « t » a donné « ć » (un « č » mouillé, mais malheureusement prononcé « č »). On voit donc bien la correspondance entre « *tele* » et « *ćelo* » et l'on prend conscience du fait que le « t » de l'augment tchèque « et » a subi lui aussi les dépalatalisations. Il est donc vraisemblable qu'il ait été originellement « et' », ce que nous considérons être une hypothèse fiable.

5 Des applications de la diachronie en synchronie

Nous attendons également des tables des langues slaves une radiographie précise de la dimension de la métathèse en haut-sorabe qui apparaît être fluctuante en comparaison avec le tchèque (figure 17). Cette connaissance issue de la comparaison entre ces deux langues est très importante.

En effet, du fait que la métathèse est totale en tchèque, cela veut dire que toutes les formes originellement « (consonne) {e/o} {l/r} consonne » sont devenues en tchèque « (consonne) {l/r} {e/a} consonne ». Il en résulte que les mots qui présentent en leur sein une structure du premier type sont étrangers au tchèque. Nous sommes actuellement en train de faire les dépouillements de vérification correspondants, afin de compléter le module de reconnaissance automatique des emprunts.

| métathèse des liquides | | | | |
|------------------------|-----|-------|-----|-------|
| haut sorabe | 0,5 | kruwa | ert | žalza |
| tchèque | 1 | kráva | ret | žláza |
| slovaque | | krava | | |
| polonais | | krowa | | |
| bas sorabe | ? | krowa | | |

Figure 17

De même, la tendance des langues slaves à ne pas débiter un mot par une voyelle doit être vérifiée quantitativement (figure 18). Nous savons qu'en tchèque à une dizaine d'exceptions près (mots outils) un mot tchèque ne peut pas commencer par « a » et à deux exceptions près par « e » et par « i ». La situation fluctuante du « i » doit être vérifiée en slovaque. Le haut-sorabe, par contre, semble être le modèle slave par excellence qui ne débute aucun mot autochtone par une voyelle. Ces connaissances sont immédiatement applicables dans la reconnaissance automatique des emprunts.

| voyelle en premier caractère d'un mot | | | | |
|---------------------------------------|-----|----|--------|----------------|
| haut sorabe | 1 | ja | a | agronom |
| | | je | e | energija |
| | | ji | i | integrować |
| | | wo | o | wokno organ |
| | | wu | u | wučbnica unija |
| tchèque | 0,6 | ja | a | apatie |
| | | je | e | embargo |
| | | ji | i | ideologie |
| | | | | okno učebnice |
| slovaque | 0,5 | i | istota | |

Figure 18

Enfin, nous venons de déposer avec nos partenaires tchèques un projet qui utilisera les données des trois tables de langues slaves dans leur dimension historique pour calculer *de facto* l'écart lexical entre langues slaves. Il s'agira de calculer en russe, slovène et langues slaves de l'Ouest des mots issus d'une même racine. La génération se fera simultanément pour toutes ces langues d'après un parcours différent suivant les phénomènes historiques concernés. Suivre les distanciations progressives entre langues slaves facilite le processus collectif de génération. Ce module pourrait être utilisé dans des systèmes de traduction automatique entre langues proches.

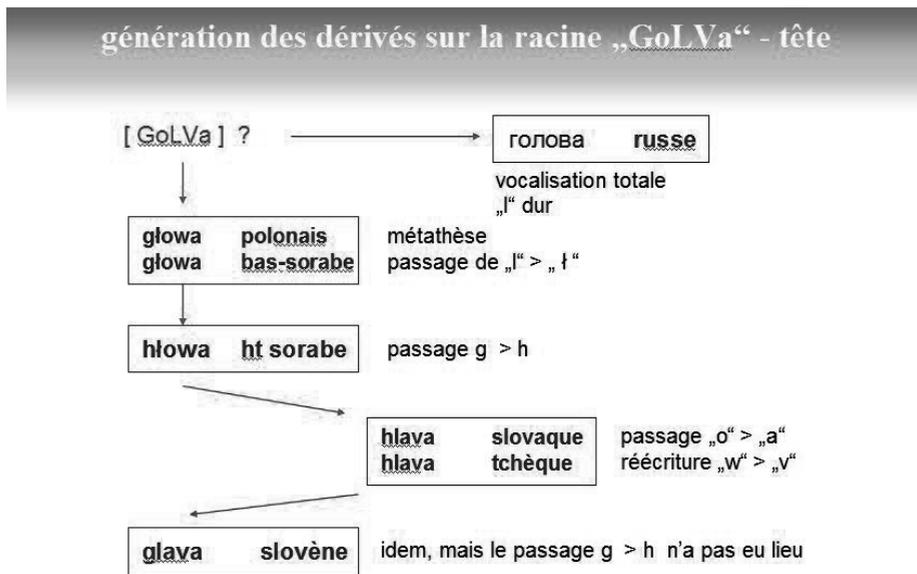


Figure 19

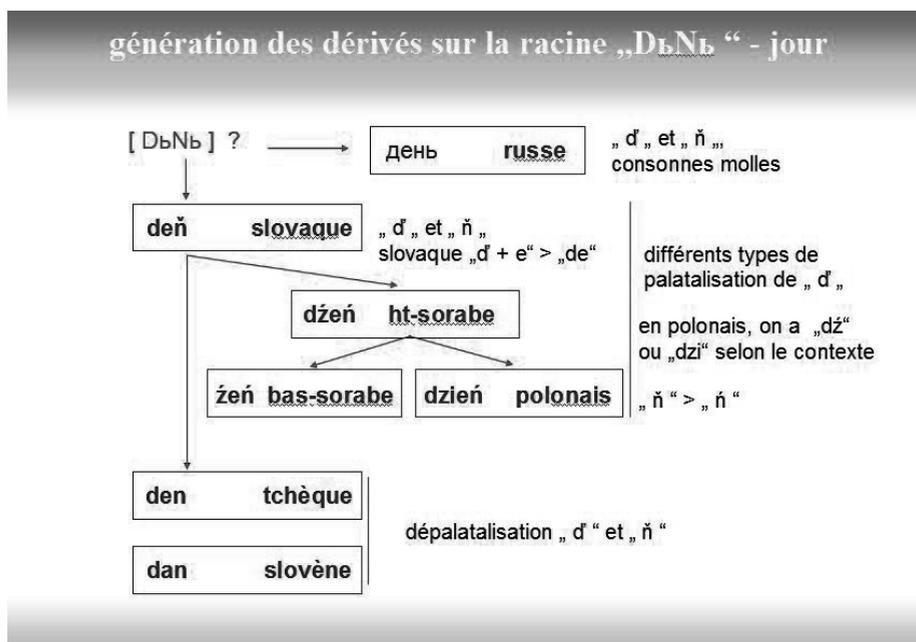


Figure 20

Les deux exemples ci-dessus concernent respectivement la génération à partir de la racine hypothétique proto-slave « *GoLVa* », *tête* qui met en jeu les phénomènes de métathèse (figure 19) et la génération à partir de la racine « *DьNь* », *jour*, qui fait appel à la vocalisation des *jers* et au processus de dépalatalisation (figure 20).

Conclusion

Nous n’avons montré ici qu’une seule des nombreuses fonctions possibles d’une base de données : l’étude du multilinguisme. Les applications de cette étude sont à trouver en premier lieu au niveau de l’enseignement et de la création d’outils d’enseignement. Le traitement automatique profite également largement de ces connaissances.

Bibliographie

- Jamborová-Lemay (D.), 2003 : *Analyse automatique du slovaque. Étude approfondie du système linguistique slovaque et sa reconnaissance d’après la forme dans les textes scientifiques et techniques. Application au machinisme agricole*. Thèse de doctorat. CERTAL-INALCO, Paris.
- Lamprecht (A.), Šlosar (D.) et Bauer (J.), 1986 : *Historická mluvnice češtiny*. Státní Pedagogické Nakladatelství, Prague.
- Lopatková (M.), Žabokrtský (Z.) et Kettnerová (V.), 2008 : *Valenční slovník českých sloves*. Karolinum, Prague.

- Mareš (F. V.), 1999 : *Diachronische Phonologie des Ur- und Frühslavischen*. Peter Lang, Frankfurt am Main.
- Marvan (J.), 2000 : *Jazykové milénium. Slovanská kontrakce a její český zdroj*. Academia, Praha.
- Mazon (A.), 1952 : *Grammaire de la langue tchèque*. Institut d'Études Slaves, Paris.
- Meškank (T.), 2009 : *Aussagenstruktur im Sorbischen. Untersuchungen zur Syntax und Satzsemantik*. Verlag Dr Kovač, Hamburg.
- Mištrík (J.), 1983 : *Moderná slovenčina*. Slovenské Pedagogické Nakladateľstvo, Bratislava.
- Pognan (P.), 1998 : « Histoire de l'écriture et de l'orthographe tchèques » dans *Histoire, Épistémologie, Langage*, Paris.
- Pognan (P.), 2001 : « Introduction aux systèmes d'écriture des langues slaves de l'Ouest (polonais, bas-sorabe, haut-sorabe, tchèque, slovaque) », dans *Slavica occitania*, Toulouse.
- Pognan (P.), 2007 : « Forme et fonction en analyse automatique du tchèque. Calculabilité des langues slaves de l'Ouest », dans *BULLAG « Les langues slaves et le français : approches formelles dans les études contrastives »*, n° 32, Presses universitaires de Franche-Comté, Besançon, p. 13-33.
- Pognan (P.), 2009a : « Définition d'une base de données linguistiques « à tout faire » comprenant le français – comme langue pivot - et les langues slaves de l'Ouest », dans *Studia Romanistica Beliana « Románske štúdie: súčasný stav a perspektívy »*, Banská Bystrica, Slovaquie, 26-27 mars 2009, p. 309-316.
- Pognan (P.), 2009b : « Définition d'un prototype général de bases de données (étude des langues slaves de l'Ouest dans une visée multilingue) » dans *MONDILEX Third Open Workshop*, Bratislava, Slovaquie, 15-16 Avril 2009, p. 183-190.
- Šewc-Schušter (H.), 1984 : *Gramatika hornjoserbskeje řeči*. Nak adništvo Domowina, Bautzen.

Mercè Vázquez García,
Antoni Oliver González,
Études des sciences de l'information et de la communication,
Arts et sciences humaines,
Universitat Oberta de Catalunya,
Barcelone,
Espagne,
mvazquezga@uoc.edu,
aoliverg@uoc.edu.

Utilisation de stratégies statistiques pour la récupération automatique de termes

Résumé

La reconnaissance automatique d'unités terminologiques et la détection précoce des néologismes sont quelques-uns des défis encore actuellement en cours lors du travail terminologique. Ces derniers constituent la base de notre recherche. D'une part, la détection automatique d'unités terminologiques et la compilation de ces unités permettent de faciliter les communications spécialisées et favorisent le travail multilingue en général ainsi que le travail des professionnels de la rédaction, de la révision, de la traduction, de la lexicographie et de la terminographie. D'autre part, l'introduction de stratégies statistiques dans le processus d'identification d'unités terminologiques rend possible le travail avec des corpus de spécialité de grand volume qui peuvent être unilingues, bilingues ou multilingues et la récupération des équivalents de la traduction.

Mots-clés: dépouillement terminologique automatisé ; mesures statistiques ; acquisition de termes ; reconnaissance de termes ; récupération d'information ; gestion des connaissances.

Introduction

Ces dernières années, le domaine de l'extraction automatique de terminologie connaît un intérêt croissant. Cette tendance est due au fait que la détection et la sélection manuelle d'unités terminologiques d'un corpus de spécialité sont des processus de travail très coûteux tant en termes de temps qu'en termes de recours à des spécialistes. L'introduction de techniques automatisées dans ce processus de travail permet de disposer des unités d'un corpus de spécialité de manière immédiate, et ainsi la tâche du spécialiste peut se concentrer sur la sélection finale des unités terminologiques.

Actuellement, en terminologie, l'acquisition et la reconnaissance des termes sont deux grands domaines de recherche très actifs liés au traitement du langage naturel. L'acquisition de termes est une tâche axée sur la localisation de nouveaux termes dans un corpus de spécialité; quant à la reconnaissance de termes, elle consiste à identifier des termes connus dans les corpus de spécialité. Cet article est consacré à la tâche de reconnaissance de termes.

Le développement de la terminologie computationnelle se répercute sur un large éventail de domaines allant de la révision, la traduction de textes spécialisés, la traduction automatisée et le résumé automatique, à la construction de glossaires et de lexiques bilingues, de vocabulaires et de dictionnaires terminologiques, l'indexation de textes, la construction de bases de données de connaissance, la construction de systèmes d'hypertexte ou de systèmes experts et l'analyse de corpus, en passant par la récupération d'informations et la gestion des connaissances.

Les techniques d'extraction de terminologie dans le domaine de la révision et de la traduction de textes spécialisés facilitent l'identification des dénominations propres à un domaine de spécialisation, la récupération automatique des équivalents de traduction qui servent à la détection de nouvelles dénominations spécialisées et permettent la publication et consultation des nouvelles dénominations dans des bases de données spécialisées. En ce qui concerne la traduction automatique et assistée, les techniques d'extraction de terminologie permettent d'élargir le lexique de spécialisation des moteurs de traduction automatique et de créer des bases de données terminologiques utiles à la traduction de textes spécialisés.

Quant à l'extraction automatique de terminologie dans le domaine de la récupération d'information, elle permet la représentation, le stockage, l'organisation et la récupération des contenus d'un domaine spécialisé; la détection des termes d'indexation ou mots clés, et permet de connaître la pertinence de certains termes par rapport à d'autres. Dans le domaine de la gestion d'information, elle permet d'accéder aux informations clés dont disposent les organisations, d'établir des réseaux de relations entre les concepts clés et de voir quels sont les nouveaux usages de l'information afin d'améliorer la gestion des connaissances, d'avoir un avantage compétitif par rapport aux autres organisations et d'accroître les connaissances au sein de l'organisation. De plus, l'utilisation de ces techniques automatisées permet une utilisation stratégique des informations car si une organisation connaît en détail son avantage compétitif par rapport à d'autres organisations, elle pourra prospérer et créer du savoir (Nonaka, Takeuchi 1995).

1 Stratégie de récupération d'unités terminologiques

L'intérêt de notre étude est axé sur la récupération du plus grand nombre d'unités terminologiques d'un corpus de spécialité et nous pensons que pour y arriver la combinaison des résultats obtenus à partir de différentes méthodes statistiques s'avère une stratégie efficace. À cet effet, d'un côté nous comparons la valeur de la

fréquence avec laquelle une unité apparaît dans un corpus de spécialité avec les résultats obtenus par treize mesures statistiques¹ afin de déterminer quelle mesure permet d'identifier plus facilement un plus grand nombre de termes. De même, nous comparons les unités qui apparaissent dans un corpus de spécialité ordonnées par fréquence avec celles qui apparaissent dans un corpus de langue générale afin de filtrer les unités qui apparaissent le plus fréquemment dans le corpus général. Cette étude présente les résultats obtenus pour le corpus *Crater*² qui appartient au domaine des télécommunications.

En premier lieu, pour déterminer la mesure statistique la plus appropriée pour identifier le plus grand nombre de termes dans un corpus de spécialité nous procédons à l'extraction automatique de candidats terme (*n-grammes*) du corpus et au filtrage de ces candidats avec une liste de mots vides, et ce, afin de disposer d'une liste d'unités ordonnées par fréquence sans inclure les variantes des candidats.

| | |
|--|--------------------------------|
| earth<>station<>1375 2249 2598 | Candidat terme et fréquence |
| signal<>unit<>958 5446 1900 | |
| signalling<>link<>866 7957 3186 | |
| data<>link<>777 3070 3186 | |
| earth<>stations<>736 2249 1441 | |
| Signalling<>System<>729 1596 976 | |
| signalling<>system<>701 7957 3294 | |
| signal<>units<>671 5446 1183 | |
| Ccitt<>Recommendation<>647 1377 2066 | |
| clear<>forward<>633 1119 881 | |
| international<>exchange<>565 3613 2524 | |
| signalling<>point<>534 7957 1720 | |
| forward<>signal<>502 1572 9373 | |
| called<>party<>469 1671 881 | |
| address<>message<>458 1902 3400 | |
| start<>stop<>399 739 549 | |
| bit<>rate<>395 1727 1442 | |
| signalling<>systems<>392 7957 2302 | |

Figure 1 : Candidats termes ordonnés par fréquence

À partir des informations sur la fréquence, nous calculons avec chaque mesure statistique les valeurs de rang et la ponctuation de chaque candidat terme sans inclure les variantes terminologiques. Les candidats, ainsi que le rang et la ponctuation, sont classés par ordre décroissant à partir de ces valeurs. Les informations relatives au rang nous indiquent la place occupée par le candidat en fonction de la ponctuation qui lui correspond pour chaque mesure statistique. Le fait que le candidat ait un rang x implique qu'il y ait $x-1$ autres ponctuations plus hautes que la ponctuation du candidat en question. Le calcul statistique de la ponctuation pour chaque candidat indique s'il y a une raison de rejeter l'hypothèse nulle ou pas, c'est-à-dire si le candidat est une collocation ou non et, en définitive, s'il peut être considéré comme un candidat susceptible d'être un terme.

1 Les treize mesures statistiques que nous utilisons sont les suivantes : le coefficient Dice, le test *Fishers twotailed*, le coefficient Jaccard, le ratio *Log-likelihood*, la mesure *True mutual information*, la mesure *Pointwise mutual information*, le ratio *Odds*, le test Chi-carré de Pearson, la mesure *T-score*, la mesure *Poisson stirling*, le coefficient PHI, le coefficient *Leftfisher* et le coefficient *Rightfisher*.

2 *CRATER, Multilingual Aligned Annotated Corpus*, www.comp.lancs.ac.uk/linguistics/Crater/corpus.html.

| | |
|---|---|
| earth<>station<>1 1.0000 1375 2249 2598 | Candidat terme, rang, position et fréquence |
| signal<>unit<>1 1.0000 958 5446 1900 | |
| signalling<>link<>1 1.0000 866 7957 3186 | |
| data<>link<>1 1.0000 777 3070 3186 | |
| earth<>stations<>1 1.0000 736 2249 1441 | |
| Signalling<>System<>1 1.0000 729 1596 976 | |
| signalling<>system<>1 1.0000 701 7957 3294 | |
| signal<>units<>1 1.0000 671 5446 1183 | |
| Ccitt<>Recommendation<>1 1.0000 647 1377 2066 | |
| clear<>forward<>1 1.0000 633 1119 881 | |
| international<>exchange<>1 1.0000 565 3613 2524 | |
| signalling<>point<>1 1.0000 534 7957 1720 | |
| forward<>signal<>1 1.0000 502 1572 9373 | |
| called<>party<>1 1.0000 469 1671 881 | |
| address<>message<>1 1.0000 458 1902 3400 | |
| start<>stop<>1 1.0000 399 739 549 | |
| bit<>rate<>1 1.0000 395 1727 1442 | |
| signalling<>systems<>1 1.0000 392 7957 2302 | |

Figure 2 : Candidats termes ordonnés par rang et ponctuation obtenus avec le coefficient Leftfisher

Ensuite, nous procédons au calcul de la valeur *tf-idf* (*term frequency-inverse document frequency*) de la liste des candidats termes obtenue au préalable à partir de notre corpus; cette valeur nous permet de comparer les candidats termes d'un corpus de spécialité à ceux d'un corpus de langue générale.

$$w_{i,j} = \frac{freq_{i,j}}{\max freq_{i,j}} \times \log \frac{N}{ni}$$

Figure 3 : Calcul de la valeur *tf-idf*

Pour le calcul de la valeur *tf-idf*, nous comparons le corpus spécialisé *Crater* au *British National Corpus*³ (BNC) anglais, de langue générale. Pour cela, nous préparons individuellement les documents de chacune des sections du corpus général, car pour calculer *tf-idf* il est important de pouvoir disposer du plus grand nombre de documents possible. De même, nous traitons le corpus spécialisé *Crater* comme un document unique car, dans les résultats finaux, nous pouvons voir quel est le poids réel de chaque candidat terme avec la valeur *tf-idf*. Si nous tenions compte des 107 documents que comprend le corpus de spécialité, les termes représentatifs seraient éparpillés et auraient un poids de *tf-idf* inférieur et peu significatif. Ensuite, nous présentons les résultats obtenus sans inclure les variantes terminologiques.

3 *British National Corpus*, www.natcorp.ox.ac.uk/.

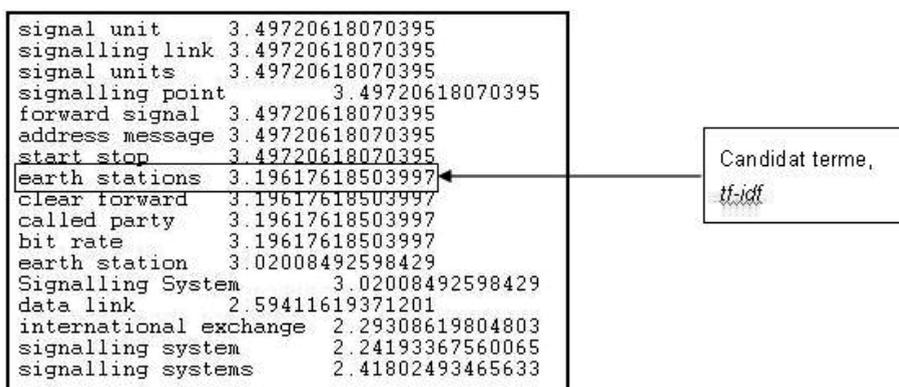


Figure 4 : Candidats termes ordonnés selon la valeur *tf-idf*

2 Évaluation des résultats obtenus

L'évaluation des résultats obtenus avec les valeurs de rang, position et *tf-idf* s'effectue à partir d'une liste de termes de référence qui appartiennent au domaine des télécommunications et nous comparons ces valeurs avec les valeurs de fréquence. Cette comparaison est nécessaire puisque la fréquence est la valeur qui indique la pertinence d'utilisation d'une dénomination déterminée dans un corpus de spécialité. Pour cette comparaison, nous classons tout d'abord la liste des termes de référence par ordre de fréquence décroissant puis nous classons les valeurs de rang, position et *tf-idf* obtenues pour chaque terme.

2.1 Évaluation des termes de référence récupérés par rang

Tout d'abord, nous évaluons les valeurs de rang des termes de référence *bi-grams* en anglais obtenues avec les treize mesures statistiques que nous utilisons. Ensuite, nous présentons un échantillon des résultats tout en indiquant quelles sont les mesures qui permettent de mieux identifier les termes de référence à partir des valeurs de rang et de fréquence.

| | TERMES REFERENCE | FREQ. | RANG | | | | | | | | | | | | |
|----|--------------------|-------|------|---------|------|-----|--------|-------|--------|-----|-------|-----|--------|-------|-------|
| | | | DICE | JACCARD | LEFT | LL | ODDS | PHI | PMI | PS | RIGHT | TMI | TSCORE | TWOT. | X2 |
| 1 | earth station | 1375 | 92 | 92 | 1 | 1 | 12.481 | 147 | 8.918 | 1 | 6.322 | 1 | 5.246 | 209 | |
| 2 | link layer | 356 | 515 | 503 | 1 | 36 | 25.354 | 845 | 11.440 | 33 | 6.322 | 28 | 19 | 5.246 | 1.708 |
| 3 | transit exchange | 315 | 627 | 608 | 1 | 51 | 28.799 | 989 | 12.822 | 47 | 6.322 | 36 | 30 | 5.246 | 2.341 |
| 4 | voice frequency | 268 | 595 | 552 | 1 | 48 | 21.763 | 838 | 9.861 | 45 | 6.322 | 35 | 46 | 5.246 | 1.674 |
| 5 | echo suppressor | 258 | 195 | 184 | 1 | 21 | 1.890 | 184 | 3.306 | 21 | 6.322 | 20 | 47 | 5.246 | 262 |
| 6 | terminal equipment | 252 | 998 | 900 | 1 | 127 | 37.246 | 1.197 | 16.798 | 124 | 6.322 | 49 | 64 | 5.246 | 4.782 |
| 7 | noise temperature | 243 | 340 | 336 | 1 | 33 | 8.413 | 431 | 5.306 | 37 | 6.322 | 28 | 56 | 5.246 | 689 |
| 8 | space segment | 218 | 172 | 171 | 1 | 31 | 5.889 | 247 | 3.431 | 34 | 6.322 | 27 | 71 | 5.246 | 364 |
| 9 | multiple access | 197 | 389 | 382 | 1 | 66 | 16.050 | 676 | 8.705 | 62 | 6.322 | 40 | 82 | 5.246 | 1.217 |
| 10 | frequency band | 180 | 983 | 899 | 1 | 167 | 31.835 | 1.172 | 13.819 | 157 | 6.322 | 52 | 109 | 5.246 | 4.240 |
| 11 | power supply | 159 | 527 | 516 | 1 | 87 | 9.654 | 626 | 5.420 | 93 | 6.322 | 43 | 123 | 5.246 | 1.102 |
| 12 | group delay | 145 | 633 | 613 | 1 | 159 | 24.043 | 1.044 | 9.534 | 148 | 6.322 | 52 | 144 | 5.246 | 2.705 |
| 13 | time slot | 144 | 644 | 621 | 1 | 99 | 6.580 | 640 | 5.204 | 105 | 6.322 | 44 | 142 | 5.246 | 1.136 |
| 14 | numbering plan | 130 | 171 | 170 | 1 | 92 | 5.760 | 283 | 2.568 | 90 | 6.322 | 43 | 169 | 5.246 | 419 |

Tableau 1 : Résultats de fréquence et valeurs de rang des termes de référence (*bi-grams* anglais)

En observant les résultats obtenus, nous constatons que deux mesures incluent par blocs de rang les termes de référence le coefficient *Leftfisher*, le coefficient *Rightfisher*, ce qui permet de localiser facilement la plupart des termes du domaine de spécialité puisqu'ils se situent dans un même rang, et que deux autres mesures, avec la valeur de rang, situent les termes de référence aux premières positions des résultats la mesure *T-score*, le test *Fishers twotailed*.

Ainsi, avec la combinaison des informations relatives à la fréquence et la valeur de rang de ces mesures, il est possible de localiser plus exactement les candidats d'un corpus de spécialité susceptibles d'être des termes. Les autres mesures ne sont pas prises en compte puisque les valeurs de rang sont réparties dans toute la liste des résultats et ne nous aident pas à distinguer les différents termes.

2.2 Évaluation de la valeur *tf-idf* des termes de référence les plus fréquents

Ensuite, nous évaluons les résultats obtenus avec les termes de référence à partir de la valeur *tf-idf* en comparaison également avec la valeur de fréquence. Nous réalisons cette évaluation pour les termes de référence *bi-grams* en anglais.

| | TERMES REFERENCE | FREQ. | TF-IDF |
|----|--------------------|-------|------------------|
| 1 | earth station | 1.375 | 3,02008492598429 |
| 2 | link layer | 358 | 3,49720618070395 |
| 3 | transit exchange | 315 | 3,49720618070395 |
| 4 | voice frequency | 268 | 3,49720618070395 |
| 5 | echo suppressor | 258 | 3,49720618070395 |
| 6 | terminal equipment | 252 | 2,79823617636794 |
| 7 | noise temperature | 243 | 3,49720618070395 |
| 8 | space segment | 218 | 3,49720618070395 |
| 9 | multiple access | 197 | 2,71905493032031 |
| 10 | frequency band | 180 | 2,35107814502572 |
| 11 | power supply | 159 | 1,63388332058350 |
| 12 | group delay | 145 | 3,49720618070395 |
| 13 | time slot | 144 | 2,38326282839712 |
| 14 | numbering plan | 130 | 3,49720618070395 |

Tableau 2 : Résultats de fréquence et *tf-idf* des termes de référence (*bi-grams* anglais)

Les résultats que nous avons obtenus avec la valeur *tf-idf* indiquent qu'il y a une correspondance claire entre la fréquence d'utilisation d'un terme et la valeur *tf-idf*, c'est-à-dire que les termes les plus fréquents ont également une valeur *tf-idf* élevée. Ainsi, une valeur *tf-idf* élevée indique qu'une unité apparaît rarement dans le corpus de langue générale et, par conséquent, qu'il y a plus de chances qu'elle soit terminologique.

2.3 Évaluation de la valeur de pondération des termes de référence les plus fréquents

Enfin, nous combinons les différents résultats obtenus (valeurs de fréquence, rang et *tf-idf*) pour les termes de référence, et nous procédons à l'évaluation conjointe par rapport à la fréquence. Cette combinaison de résultats correspond à ce que

nous appelons *valeur de pondération*. Les résultats de rang correspondent aux *bi-grams* du coefficient *Leftfisher*.

| | TERMES RÉFÉRENCE | FREQ. | PONDERATION |
|----|--------------------|-------|-------------------|
| 1 | earth station | 1.375 | 0.95452357967414 |
| 2 | link layer | 356 | 0.752969696969697 |
| 3 | transit exchange | 315 | 0.743030303030303 |
| 4 | voice frequency | 268 | 0.731636363636364 |
| 5 | echo suppressor | 258 | 0.729212121212121 |
| 6 | terminal equipment | 252 | 0.661135824163154 |
| 7 | noise temperature | 243 | 0.725575757575758 |
| 8 | space segment | 218 | 0.719515151515152 |
| 9 | multiple access | 197 | 0.640255395428876 |
| 10 | frequency band | 180 | 0.601060778397108 |
| 11 | power supply | 159 | 0.527611033959345 |
| 12 | group delay | 145 | 0.701818181818182 |
| 13 | time slot | 144 | 0.595401162318065 |
| 14 | numbering plan | 130 | 0.698181818181818 |

Tableau 3 : Résultat de fréquence et pondération des termes de référence (bi-grams anglais)

Nous observons que la valeur de taux de pondération facilite l'identification des termes d'un corpus de spécialité puisqu'elle est en corrélation avec la valeur de fréquence d'utilisation des termes: les valeurs élevées de fréquence des termes de référence dans le corpus ont également une valeur de pondération élevée.

3 Comparaison des résultats en incluant les variantes terminologiques

Nous avons obtenu les résultats présentés jusqu'à maintenant en évaluant séparément toutes les variantes terminologiques des termes de référence figurant dans le corpus de spécialité. C'est pourquoi nous avons voulu introduire dans l'analyse des résultats l'ajout des différentes variantes terminologiques singulier, pluriel et majuscule et minuscule afin de vérifier si cela permet d'obtenir des résultats plus exhaustifs. En prenant cet aspect en compte, nous avons tout d'abord ajouté les variantes terminologiques du corpus de spécialité et, à partir de là, nous avons récupéré les candidats termes classés par fréquence afin de disposer des valeurs de rang et de position de ces candidats.

Les résultats que nous avons obtenus à partir de l'ajout des différentes variantes terminologiques nous permettent d'avoir, d'une part, la fréquence totale des termes de référence dans le corpus, que nous obtenions partiellement jusqu'alors. D'autre part, si nous comparons les résultats de rang ajoutés avec ceux qui ne sont pas ajoutés, parmi les quatre mesures qui offrent la meilleure corrélation entre fréquence et rang, nous observons que le coefficient *Leftfisher* continue de placer les termes les plus fréquents dans le rang 1, le coefficient *Rightfisher* les place dans un rang très inférieur (du 6 322 au 1 732), la mesure *True mutual information* met les termes de référence dans des rangs inférieurs à ceux de la mesure *T-score* et le test *Fishers twotailed* les met également dans un rang inférieur (du 5 246 au 1 682). Ci-après, le tableau complet des résultats de fréquence et de rang de toutes les mesures.

| | TERMES REFERENCE | FREQ. | RANG | | | | | | | | | | | | |
|----|--------------------|-------|------|-----|---------|------|-------|------|-------|-----|-------|-----|-------|-------|------|
| | | | DICE | LL | JACCARD | LEFT | ODDS | PHI | PMI | PS | RIGHT | TMI | TSORE | TWOT. | X2 |
| 1 | earth station | 2173 | 61 | 1 | 61 | 1 | 3693 | 67 | 7950 | 1 | 1732 | 1 | 1 | 1682 | 89 |
| 2 | link layer | 471 | 74 | 8 | 74 | 1 | 91 | 74 | 3118 | 8 | 1732 | 6 | 16 | 1682 | 96 |
| 3 | transit exchange | 372 | 514 | 38 | 503 | 1 | 20794 | 814 | 10053 | 35 | 1732 | 13 | 31 | 1682 | 1531 |
| 4 | voice frequency | 362 | 534 | 39 | 521 | 1 | 21003 | 835 | 10159 | 38 | 1732 | 13 | 35 | 1682 | 1610 |
| 5 | echo suppressor | 358 | 635 | 44 | 616 | 1 | 24038 | 971 | 11730 | 43 | 1732 | 14 | 36 | 1682 | 2142 |
| 6 | terminal equipment | 302 | 597 | 48 | 582 | 1 | 18587 | 823 | 9199 | 45 | 1732 | 14 | 60 | 1682 | 1574 |
| 7 | noise temperature | 298 | 876 | 81 | 825 | 1 | 28728 | 1151 | 14179 | 78 | 1732 | 15 | 61 | 1682 | 3607 |
| 8 | space segment | 280 | 274 | 36 | 274 | 1 | 5048 | 311 | 4419 | 39 | 1732 | 13 | 70 | 1682 | 467 |
| 9 | multiple access | 235 | 384 | 50 | 382 | 1 | 5951 | 417 | 4747 | 59 | 1732 | 14 | 95 | 1682 | 666 |
| 10 | frequency band | 234 | 342 | 67 | 342 | 1 | 13115 | 581 | 5975 | 65 | 1732 | 15 | 96 | 1682 | 959 |
| 11 | power supply | 226 | 172 | 118 | 172 | 1 | 5472 | 250 | 3321 | 48 | 1732 | 14 | 99 | 1682 | 365 |
| 12 | group delay | 226 | 800 | 47 | 766 | 1 | 25602 | 1123 | 12061 | 114 | 1732 | 17 | 100 | 1682 | 3278 |
| 13 | time slot | 195 | 447 | 78 | 442 | 1 | 6712 | 489 | 4706 | 85 | 1732 | 15 | 131 | 1682 | 791 |
| 14 | numbering plan | 192 | 797 | 143 | 764 | 1 | 23943 | 1115 | 11069 | 137 | 1732 | 17 | 137 | 1682 | 3182 |

Tableau 4 : Résultats de fréquence et rang des termes de référence ajoutés (bi-grams anglais)

Pour pouvoir déterminer, parmi les quatre mesures, celle qui récupère le plus grand nombre de termes à partir de l'information de rang, nous avons calculé le pourcentage de termes de référence en tenant compte du nombre de candidats termes de chaque rang. Ainsi, dans le rang 1 du coefficient *Leftfisher* on trouve 407 des 432 termes de référence du corpus. En tenant compte de cette seule donnée, nous pouvons dire que c'est le coefficient qui récupère un plus grand nombre de termes. Cependant, c'est dans ce même rang 1, que se trouvent la plupart des candidats termes et nous ne pouvons donc pas le considérer comme aussi efficace pour localiser les termes. En revanche, la mesure *True mutual information*, qui répartit les termes dans plusieurs rangs, a un pourcentage élevé de termes dans chaque rang, si nous tenons compte du nombre total de candidats termes se trouvant également dans chaque rang. Nous en déduisons donc que, des quatre mesures, la mesure *True mutual information* est celle qui permet de mieux localiser les termes de référence à partir de la valeur de rang. Le tableau suivant montre les résultats obtenus.

| MESURES | RANGS | TERMES REFERENCE PAR RANG | CANDIDATS TERMES PAR RANG | POURCENTATGE TERMES REFERENCE PAR RANG |
|-----------|-----------|---------------------------|---------------------------|--|
| LEFT | Rang 1 | 407 / 432 | 60.479 / 66.312 | 0,67% |
| RIGHT | Rang 1732 | 373 / 432 | 47.579 / 66.312 | 0,78% |
| TMI | Rang 1 | 1 / 432 | 1 / 66.312 | 100% |
| | Rang 6 | 1 / 432 | 2 / 66.312 | 50% |
| | Rang 13 | 3 / 432 | 7 / 66.312 | 42,8% |
| | Rang 14 | 4 / 432 | 15 / 66.312 | 26,6% |
| | Rang 15 | 2 / 432 | 24 / 66.312 | 8,3% |
| | Rang 16 | 3 / 432 | 35 / 66.312 | 8,57% |
| Rang 17 | 5 / 432 | 59 / 66.312 | 8,47% | |
| TWOTAILED | Rang 1682 | 375 / 432 | 47.580 / 66.312 | 0,79% |

Tableau 5 : Pourcentage de termes de référence par rang

Conclusion

En conclusion, la combinaison de plusieurs stratégies statistiques au moyen de la valeur de pondération facilite l'identification des unités terminologiques de référence d'un corpus de spécialité. En outre, avec l'ajout de variantes terminologiques,

il est possible de récupérer de façon plus exhaustive les unités terminologiques ainsi que leur fréquence d'utilisation dans le corpus. Pour finir, l'utilisation de certaines mesures statistiques facilite la détection automatique des unités terminologiques d'un corpus de spécialité.

Bibliographie

- Banerjee (S) et Pedersen (T.), 2003 : « The Design, Implementation and Use of the Ngram Statistics Package », dans *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, México, dernière mise à jour le 23 octobre 2009, www.d.umn.edu/~tpederse/Pubs/cicling2003-2.pdf.
- Bourigault (D.), 1994 : *LEXTER, un logiciel d'Extractions de Terminologie. Application à l'acquisition des connaissances à partir des textes*, École des Hautes Études en Sciences Sociales.
- Bourigault (D.), Gonzalez-Mullier (I.) et Gros (C.), 1996 : « LEXTER, a Natural Language Processing Tool for Terminology Extraction », dans *Proceedings of the VII EURALEX International Congress on Lexicography* (EURALEX 1996), p. 771-779.
- Bourigault (D.), Jacquemin (C.) et L'Homme (M.-Cl.), éd., 2001 : *Recent Advances in Computational Terminology*, Amsterdam, John Benjamins.
- Brookes (B. C.), 1980 : « The foundations of Information Science », dans *Journal of Information Science*, vol. 2, n° 3-4, p. 125-33.
- Brookes (B. C.), 1981 : « The foundations of information science », dans *Journal of Information Science*, vol. 3, p. 3-12.
- Cabré (M.T.), 1992 : *La terminologia. La teoria, els mètodes, les aplicacions*, Barcelone, Empúries.
- Cabré (M.T.), 1999 : « Una nueva teoría de la terminología: de la denominación a la comunicación », dans *La terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*, Barcelone, Institut Universitari de Lingüística Aplicada, UPF.
- Codina (L.) et Rovira (C.), 2002 : « Information Retrieval Techniques », dans *Organización y recuperación de la información*, Documents de lectura, Universitat Oberta de Catalunya, dernière mise à jour le 23 octobre 2009, cv.uoc.es/cdocent/BOIQM7V2N6_PVI7JZGVG.pdf.
- Daille (B.), 1994 : « Combined approach for terminology extraction: lexical statistics and linguistic filtering », dans *UCREL*, n° 5, Université de Lancaster, dernière mise à jour le 23 octobre 2009, http://ucrel.lancs.ac.uk/tech_papers.html.
- Daille (B.), Gaussier (E.) et Lang (J.), 1998 : « An Evaluation of Statistical Scores for Word Association », dans *Proceedings of the Tbilisi Symposium on Logic, Language and Computation*, CSLI Publications, p. 177-188.
- Jacquemin (C.) et Bourigault (D.), 2003 : « Term-Oriented NLP », dans *The Oxford Handbook of Computational Linguistics*, éd. Ruslan Mitkov, Oxford University Press.
- Kageura (K.) et Umino (B.), 1996 : « Methods of Automatic Term Recognition: a review », dans *Terminology*, vol. 3, n° 2, p. 259-289, dernière mise à jour le 23 octobre 2009, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.25.3740>.
- Nakagawa (H.) et Mori (T.), 2000 : « Automatic Term Recognition based on Statistics of Compound Nouns », dans *Terminology*, vol. 6, n° 2, p. 195-210, dernière mise à jour le 23 octobre 2009, www.r.dl.itc.u-tokyo.ac.jp/~nakagawa/academic-res/term9-2.pdf.

Nonaka (I.) et Takeuchi (H.), 1995 : *The knowledge-creating company: How Japanese Companies Create the Dynamics of Innovation*, Nova York, Oxford University Press.

Universitat Politècnica de Catalunya; *TERMCAT*, Centre de Terminologia et Enciclopèdia Catalana, 2007 : *Diccionari de telecomunicacions*, Barcelona, Enciclopèdia Catalana/TERMCAT, Centre de Terminologia, ISBN 978-84-412-1459-0 ; 978-84-393-7408-4.

Vivaldi (J.), Màrquez (L.) et Rodríguez (H.), 2001 : « Improving Term Extraction by System Combination Using Boosting », dans *Lecture Notes in Computer Science*, n°. 2167, p. 515-526, dernière mise à jour le 23 octobre 2009, citeseer.ist.psu.edu/vivaldioimproving.html.

Isabelle Oliveira,
Université Paris 3, Sorbonne Nouvelle,
Paris,
France,
ioliveira@isc.cnrs.fr.

Sabine Ploux,
L2C2, Institut des sciences cognitives, UMR 5 230 CNRS-Université Lyon 1,
Lyon,
France,
sploux@isc.cnrs.fr.

Vers une méthode de détection et de traitement automatique de la métaphore

Résumé

Nous présentons un modèle pour le repérage automatique de la métaphore et pour l'aide à son analyse. Les résultats actuels de ce modèle sont très prometteurs. La présentation du modèle est suivie d'une analyse comparative de la métaphore en langue de spécialité (ici le domaine du droit) et en littérature. Nous projetons de confronter ces deux approches afin à la fois de valider systématiquement le système d'analyse automatique et de donner une base objective à l'étude linguistique de la métaphore.

Mots-clés: métaphore; traitement automatique des langues; corpus.

Introduction

La métaphore terminologique, processus linguistique fondé sur le rapprochement mental qu'opère l'utilisateur d'une langue entre deux objets ou phénomènes par association d'idées, est une arme précieuse dans l'arsenal linguistique du juriste, mais aussi du littéraire. Cependant, en langue de spécialité on cherche à se démarquer de la métaphore qui renvoie souvent à la polysémie et à l'ambiguïté. En effet, cette figure de style consistant en un transport de sens risque à tout moment de rendre inopérant l'effort d'univocité et de rigueur, mais la métaphore n'est pourtant pas absente des langues de spécialité. Il faut alors se demander si cette présence est accidentelle, contingente, ou si elle est incontournable et signifiante. Afin d'étudier ces questions, nous avons cherché dans un premier temps à nous donner les moyens de procéder à une analyse systématique effectuée sur des corpus de textes. Dans la première partie de l'article, nous présenterons, l'ensemble des

corpus constitués et les méthodes de repérage et d'analyse des métaphores qui en sont extraites. Dans la seconde partie, nous présentons une analyse linguistique comparative de la métaphore littéraire et de la métaphore dans une langue de spécialité: le droit. Notre travail est prospectif et nous projetons dans l'avenir de rapprocher et de faire interagir ces deux facettes de l'analyse afin à la fois d'affiner la démarche automatique et de nourrir l'étude linguistique de résultats obtenus non plus « à la main », mais par un traitement systématique et automatique.

1 Une approche fondée sur l'analyse de corpus et le traitement automatique des langues

1.1 Description des corpus

Les corpus que nous avons compilés ont été regroupés suivant leur genre: littéraire, journalistique et de droit. Du point de vue de l'étude de la métaphore, ces trois genres présentent des caractéristiques différentes qui, comme nous le montrerons dans la seconde partie de l'article, justifient leur choix. Chaque corpus contient des textes en français et des textes en portugais. Nous nous sommes efforcées, quand cela était possible, de constituer des corpus parallèles et sinon comparables. Ce choix permettra par la suite une étude translingue du phénomène métaphorique.

- **Corpus littéraire français-portugais:** le corpus littéraire contient environ 200 romans du XIX^e siècle français ou portugais et leur traduction dans l'autre langue. Il a été constitué grâce aux textes disponibles sur les sites des projets *Gallica* de la Bibliothèque nationale¹ et *Gutenberg*² et par numérisation des traductions ou des textes en langue originale correspondants. Le travail de numérisation a été effectué grâce à la plate-forme technologique de la Maison de l'Orient et de la Méditerranée de l'Université Lyon 2³.
- **Corpus journalistique:** le corpus journalistique contient cinq années du journal *Le Monde* (1997-2001). Nous procédons actuellement à la constitution d'un corpus comparable à partir de quotidiens portugais.
- **Corpus de droit:** le corpus parallèle de droit contient les traités communautaires (Constitution européenne, Traités d'Amsterdam, de Lisbonne, de Maastricht, de Madrid, de Rome et le Traité constituant la Communauté européenne de l'énergie atomique (CEEA)) dans les deux langues française et portugaise.

1.2 Vers une détection automatique de la métaphore

Le repérage automatique ou semi-automatique des métaphores constitue sans doute à la fois un défi majeur pour le traitement automatique des langues et une perspective utile pour le travail linguistique. Dans ce projet, nous avons cherché un critère utile à ce repérage. Ce critère est fondé sur la définition même de la métaphore, à savoir l'association entre des domaines source et cible hétérogènes et repose sur

1 BNF, gallica.bnf.fr/.

2 www.gutenberg.org.

3 Patrick Desfarges, responsable.

une mesure de la distance sémantique entre les domaines en interaction. Ainsi, dans la métaphore juridique *fruit civil*, les domaines hétérogènes de la botanique et du droit sont liés et associés pour produire le sens voulu. *A contrario*, dans l'expression *fruits et légumes*, les deux termes (*fruit* et *légume*) ont des contextes d'emploi très souvent similaires. Dans ce travail, nous proposons d'utiliser le modèle ACOM (*Automatic Contexonym Organizing Model*) (Ji *et al.* 2003) des *Atlas sémantiques* (AS) pour calculer une distance entre les contextes d'emploi les plus génériques des termes d'une expression (métaphorique ou non) et nous utilisons le résultat de ce calcul comme critère de détection d'une métaphore.

Le modèle ACOM, dont on peut consulter les résultats à l'adresse dico.isc.cnrs.fr, fournit pour chaque mot une liste de mots associés qui représentent et caractérisent son emploi. Il est fondé sur une analyse systématique de corpus et a été testé pour la valeur cognitive de ses résultats. En effet, les mots proposés par ACOM en réponse à une requête sont souvent identiques aux mots donnés par des sujets humains lors d'une tâche d'association (Ji *et al.* 2008). Ainsi, les premiers mots associés au mot *poire* et donnés par le modèle sont *abricot, chou-fleur, couper, morceaux, doigts, fromage, fruits, pêche...*

On pourra se reporter à l'article mentionné ci-dessus pour plus de précisions concernant le modèle et sa mise en œuvre. Dans la présente étude, nous l'avons utilisé de la façon suivante. Nous avons voulu mesurer le recouvrement thématique entre les termes d'une expression métaphorique par un calcul du recouvrement entre les mots associés à chacun d'eux. L'hypothèse sous-jacente consiste à prédire que dans une métaphore le recouvrement sera réduit car les domaines source et cible sont par définition différents alors que dans une expression non métaphorique ce recouvrement devrait être plus important car les mots associés sont au moins partiellement partagés.

1.3 Description de l'expérience réalisée sur corpus

Afin de tester cette hypothèse nous avons constitué des bases de syntagmes nominaux et verbaux extraits du corpus littéraire français et du corpus journalistique français pour en extraire des expressions porteuses de métaphores et des ensembles contrôle d'expressions non métaphoriques.

1.3.1 Constitution des bases de syntagmes à partir des corpus

La méthode de constitution de ces bases est décrite dans (Ben Maaouia 2009). Elle utilise en prétraitement une analyse morphologique des corpus effectuée grâce à l'analyseur *TreeTagger*⁴ de l'Université de Stuttgart, disponible pour de nombreuses langues, dont le français et le portugais. Ces bases constituées contiennent l'ensemble des syntagmes nominaux ou verbaux contenus dans les corpus. Pour chaque syntagme sont également donnés : sa fréquence d'emploi dans le corpus et l'ensemble des lemmes et des mots de fonction qu'il contient.

4 www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

1.3.2 Sélection d'expressions métaphoriques et d'expressions non métaphoriques appariées en fréquence d'emploi

De chacune de ces trois bases (base de syntagmes nominaux extraits du corpus littéraire, base de syntagmes verbaux extraits du même corpus et base de syntagmes nominaux extraits du corpus journalistique), deux ensembles ont été sélectionnés. Le premier ensemble contient des expressions métaphoriques sélectionnées et validés par I. Oliveira. Le second ensemble a été obtenu par extraction automatique à partir de chacune des 3 bases et de façon aléatoire d'un groupe d'expressions non métaphoriques. Nous avons ensuite vérifié que chacune des paires de groupes (les expressions métaphoriques et non métaphoriques de chacun des corpus) sont appariées en fréquence d'emploi.

1.3.3 Mesure du recouvrement thématique

Enfin, pour chaque expression sélectionnée et pour chaque mot plein de cette expression nous avons calculé les cinquante premiers mots associés donnés par le modèle ACOM⁵. Le programme que nous avons réalisé calcule ensuite le nombre de mots associés communs à l'ensemble des mots pleins de l'expression donnée⁶. Ainsi pour l'expression non métaphorique *un visage rose* l'ensemble de mots associés communs au mot *visage* et au mot *rose* comportent les mots *blanc, rouge, main, fille, visage, rose* et la mesure vaut donc 6. Nous avons ensuite procédé à une analyse statistique (t-test pour échantillons indépendants) comparative des résultats entre expressions métaphoriques et expressions non métaphoriques appariées en fréquence d'emploi. Les résultats sont significatifs pour chacune des bases et montrent donc un recouvrement entre mots associés nettement plus faible pour les expressions métaphoriques que pour les expressions non métaphoriques. La figure 1 donne les résultats pour les expressions nominales extraites du corpus littéraire. Nous donnons ici les valeurs du test sur les autres séries : (1) Moyennes du nombre de mots associés communs pour les expressions verbales métaphoriques $M = 0,806763$ et non métaphoriques $M = 2,464213$ du corpus littéraire (t-test (299) = -5,75359, $p < 0,001$), (2) Moyenne du nombre de mots associés communs pour les expressions nominales métaphoriques $M = 4,507576$ et non métaphoriques $M = 8,871795$ du corpus journalistique (t-test (81) = 8,871795, $p < 0,001$).

⁵ Les paramètres α et β du modèle qui fixent la proportion des mots cooccurrents du mot initial retenu sont fixés à 0.2.

⁶ Si l'expression contient plus de deux mots pleins, le programme calcule la moyenne deux à deux du nombre de mots associés communs aux mots pleins de l'expression considérée.

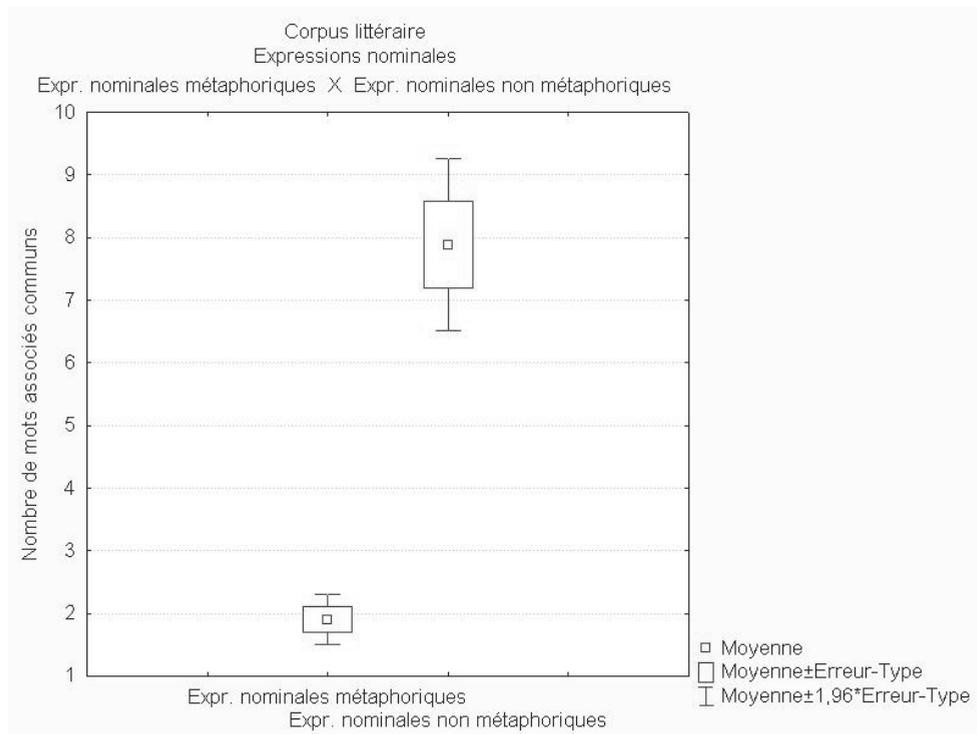


Figure 1: Moyennes du nombre de mots associés communs pour les expressions nominales métaphoriques $M = 1,903704$ et non métaphoriques $M = 7,884615$ du corpus littéraire. $t\text{-test}(211) = -10.08$, $p < 0,001$

Cette mesure du recouvrement thématique semble donc être un critère très prometteur pour le repérage automatique des métaphores.

1.4 Vers une aide à l'analyse automatique de la métaphore

Outre le repérage, nous cherchons une méthode automatique pour l'aide à la détermination des éléments qui participent au processus métaphorique. Un premier exemple de cette démarche a été présenté en 2008 (Ploux et Oliveira 2008). En effet, dans l'analyse d'une métaphore, la mise en évidence des traits sémantiques opérants est un travail d'experts dont le résultat varie suivant les auteurs. Il est alors intéressant de chercher une méthode objective qui ne repose pas uniquement sur des interprétations, interprétations qui ne sont ni toujours congruentes ni toujours exhaustives. Notre ambition est alors de prendre support sur une analyse systématique de corpus pour l'élaboration d'un modèle capable de mettre en évidence automatiquement des traits sémantiques à l'œuvre dans une métaphore. La proposition encore embryonnaire que nous faisons ici repose comme précédemment sur le modèle géométrique ACOM. Ji et ses collègues (Ji *et al.* 2003) ont montré que ce modèle permet, outre l'utilisation qui en a été décrite au paragraphe précédent, de calculer, pour une liste de mots donnée, des termes qui représentent les contextes d'emploi qui lient les mots de la liste. À titre d'exemple, nous donnons ici

pour quelques métaphores (1) les premiers termes associés à chacun des termes de l'expression métaphorique pris isolément ainsi que (2) les premiers termes associés à l'ensemble de la métaphore. On voit sur ces exemples se dégager quelques traits sémantiques qui participent à la cohésion métaphorique.

– *Une foule tiède*

- foule : arrivée, mille, feu, milliers, bras, capitale
- tiède : eau, robinet
- foule + tiède : bain, feu, fête, images, immense

Ici, le mot *bain* commun à la fois aux expressions courantes *bain de foule* et *bain tiède* est détecté par le modèle comme un élément porteur des images qui émergent de l'expression métaphorique *foule tiède*.

– *La bataille de l'argent :*

- bataille : champ, issue, perdu, cheval, livrer
- argent : armes, marchés, permet, problème, utilisation
- bataille + argent : perdu, perdre, lutte, gagner, gagné.

Ici les mots *perdre*, *gagner* et *lutte* sont proposés par le modèle comme des éléments porteurs de la cohésion sémantique de l'expression *la bataille de l'argent* et représentent effectivement ce qu'argent et bataille partagent au-delà de leurs spécificités propres.

Nous voulons dans le futur réaliser une expérience en grandeur nature afin de confronter les résultats du modèle à ceux donnés par une analyse linguistique et tester à la fois sa validité et son utilité pour les linguistes.

La suite de l'article présente une analyse linguistique faite en parallèle à ce travail de modélisation et qui est donc un élément essentiel de cette confrontation programmée.

2 Analyse linguistique comparative

2.1 La terminologie littéraire : le perspectivisme

De notre corpus littéraire se profile non seulement la reconnaissance de la spécificité des langues littéraires dans leur diversité, mais aussi la reconnaissance de la spécificité de l'usage littéraire du langage. Un langage littéraire qui exprime le génie individuel de son auteur. Dans ce sens, la vérité dépend largement de la perspective choisie du réel. Il n'y a pas de réel en soi, pas de vérité absolue, pas de sens unique et fondamental qui épuiserait les significations de l'être. Le choix de la perspective va dépendre de la sélection minutieuse des valeurs sémantiques que le lecteur aura privilégiées dans son interprétation, dans sa perspective. Aucune

perspective n'est purement logique, c'est-à-dire neutre, objective. Le discours littéraire se distingue ainsi par la richesse de ses images et le jeu sur les dénominations. En littérature, cette figure de rhétorique consistant en une translation, un transfert de sens risque à tout moment de rendre inopérant l'effort d'univocité et de rigueur. D'abord, la métaphore permet de désigner les opérations jusque-là inaperçues de la conscience : la conscience « vise », « saisit », « suspend ». Toutefois, lorsqu'il s'agit de désigner la subjectivité absolue qui n'est aucunement constituée, la métaphore reste le seul recours possible. Cette subjectivité absolue est « quelque chose qu'il faut désigner métaphoriquement comme « flux », quelque chose qui jaillit « maintenant », en un point d'actualité, un point source originale... » Le discours littéraire fait ici l'épreuve de ses limites, non parce qu'il n'est pas assez riche, mais parce que la subjectivité absolue n'est aucunement constituée et échappe au langage qui ne peut désigner que le constitué. Le constitutif lui échappe par nature. La métaphore ne devient véritablement intéressante que lorsqu'elle ne s'adapte pas parfaitement à ce qu'elle entend désigner et oblige parfois l'écrivain à commenter et interpréter le dispositif qu'il a lui-même élaboré par le biais d'une métaphore filée. Il prend appui sur elle en même temps qu'il en marque les limites et parvient ainsi à nous amener à une compréhension de la réalité qu'il désigne.

La métaphore est aussi le lien d'une thématique de transfert qui dynamise les représentations construites par les textes. Déplacements corporels, métamorphoses, animation des inanimés de façon pittoresque, sont autant de thèmes récurrents qui constituent l'univers littéraire. La métaphore en littérature lie le sensible et l'intelligible. En voyant ou en lisant, on comprend. Elle n'est rien d'autre que ce miroir qui nous anime par le mouvement, ayant pour caractéristique principale d'être créative. Il y aurait ainsi une dynamique propre à la métaphore littéraire qui inviterait à poursuivre l'analogie ou à la critiquer, c'est-à-dire à penser puisque nous voyons ici en quel sens l'on peut dire que pensée et langage ne sont pas deux temps séparés, mais les deux faces d'un même processus de recherche.

2.2 Spécificité de la langue juridique

La langue juridique s'associe plus à la logique, qui vise le repérage d'entités absolument stables, non affectées par l'incessante ambigüité, ambivalence, qui apparaissent ou disparaissent ou se modifient sans répit. Tout l'édifice conceptuel, rationnel qu'est le texte juridique serait donc miné en son fondement par la métaphore. Mais une telle approche n'est-elle pas elle aussi exposée à la critique ? En effet, si les concepts sont originellement métaphoriques, est-ce que le travail du juriste ne consiste justement pas à limiter le glissement incessant des dénominations et des concepts pour parvenir à une détermination rationnelle ?

Le discours juridique s'efforce de travailler à partir de concepts précis dont le sens doit être délimité afin de bannir les ambigüités et les fausses évidences qu'elles véhiculent. La démarche du juriste consiste alors à rapporter l'inconnu au connu, ce qui revient finalement à faire usage d'une métaphore, à désigner l'inédit par une dénomination déjà connue et à déplacer ainsi le sens attaché à cette dénomination,

à le transporter. Mais le problème est que le discours rationnel finit par prendre des métaphores pour des essences. Dans *La Métaphore vive*, dans sa huitième étude intitulée « Métaphore et discours philosophique », Paul Ricœur souligne ainsi que ce n'est pas parce que la métaphore est à l'origine du concept qu'elle continue d'agir en lui. Si l'empreinte métaphorique est effacée sur la monnaie conceptuelle, alors il n'y a plus de glissement possible et d'effet inaperçu. Si le texte juridique est riche en métaphores « lexicalisées », une fois que cette opération de lexicalisation a eu lieu, une signification propre est produite qui devient indépendante par rapport au premier sens du mot. La question du rapport entre métaphore et texte juridique se noue finalement autour de l'idée d'analogie. Mais si ce jeu sur le semblable est possible, c'est parce qu'il y a du même, parce que l'identité fonde la ressemblance. Si ressemblance il y a entre diverses réalités, n'est-ce pas parce qu'on peut les référer en dernière instance à l'Idée ? Si nous reprenons l'image de *abeillage* dans le droit seigneurial, si ces différentes abeilles ont une ressemblance entre elles, c'est bien parce qu'elles ont à voir avec l'Idée d'abeille, avec une unité et une identité qui rendent raison de cette ressemblance. Si le « même » fonde le « semblable », si le concept est logiquement premier par rapport à la métaphore, celle-ci est peut-être déterminante dans l'ordre de la découverte. C'est ce que Ricœur appelle la « métaphore vive » ou dans notre optique une métaphore « in vivo ». Ce qui résulte de l'irruption d'une métaphore, ce n'est pas un savoir, un donné, mais une demande en concept. Cet écart sémantique en lequel consiste la métaphore sollicite la pensée. L'interprétation est à l'intersection de ces deux mouvances en ce qu'elle recherche la clarté du concept tout en se nourrissant de la force et de la richesse de la métaphore « in vivo » qui relance la réflexion quand le concept l'a figée. Si la métaphore a une place décisive dans le texte juridique, c'est bien parce que la pensée spéculative se sait différente et capable d'assumer le défi que lui lance la métaphore qui demande toujours à être travaillée par le juriste en même temps qu'elle lui est nécessaire. Si le discours juridique a besoin des possibilités offertes par la métaphore, la rigueur du concept permet seule d'en éviter les effets indésirables.

Néanmoins, si la métaphore apparaît comme irréductible, en fait, son danger le plus grand n'est pas sa présence, mais le fait qu'on oublie sa présence, qu'elle s'efface comme métaphore et agit alors sans que l'on en puisse maîtriser les effets. Elle tend à identifier des registres hétérogènes et à introduire une confusion périlleuse. Ainsi, le mot *fortune* dans l'expression *fortune de mer* ne se réfère pas à la richesse. Le mot est pris dans son sens latin d'« aléa ». Il s'agit d'un événement fortuit qui peut être heureux ou malheureux. Ici il est pris dans le sens d'événement dommageable qui se produit au cours d'une expédition maritime, tel un naufrage. C'est donc un cas de force majeure propre au droit maritime.

Si les dangers de la métaphore en langue de spécialité sont réels, ne peut-on pas envisager un usage maîtrisé de celle-ci qui la subordonnerait au concept tout en servant sa cause. En ce sens, on peut mettre en évidence des fonctions de la métaphore dans le texte juridique. Elle a d'abord un rôle didactique en ce qu'elle permet de présenter de façon imagée, en se référant au sensible, des relations abstraites. La métaphore terminologique est aussi un moyen de tisser des liens entre

des domaines apparemment distincts et de révéler la systématique et les analogies présentes dans la réalité elle-même. Elle est un outil particulièrement efficace pour retrouver des identités de structure au-delà de la multiplicité des domaines concernés. La métaphore terminologique, enfin, permet au juriste de suppléer aux insuffisances du langage, lequel ne peut offrir de nouvelles dénominations qui correspondraient à une façon inédite d'appréhender la réalité. La démarche phénoménologique est intéressante à cet égard. Elle fait l'épreuve des limites du langage naturel une fois l'époque opérée.

Ainsi, souvent, ce que l'on essaie de dire « n'existe jamais réellement », c'est pourquoi nous percevons la métaphore terminologique comme le réceptacle et pour ainsi dire la nourrice de tout ce qui naît. Pourtant, notre incapacité d'abstraire totalement, la finitude, la rendent omniprésente. La déficience est alors du côté du sujet parlant et écrivant qui a besoin de l'appui des images pour penser.

2.3 Le problème de l'équivalence en littérature et dans le domaine juridique

Le discours littéraire reprend souvent le modèle du vivant, faisant appel dans sa démarche de structuration à des métaphores de l'organisme ou du corps.

Exemples en littérature :

- Le sang crevant sa face large = *O sangue a saltar-lhe das faces rubicundas*
- La poussière s'obstinait là = *O pó teimava em se juntar*
- Le soleil peuplait les bancs des poussières qui dansaient dans ses rayons = *O sol povoava os bancos com a poalha que dançava nos seus raios*
- Une foule tiède = *Uma multidão calorosa*
- Les herbes qui soupiraient d'aise = *As ervas que suspiravam de contentamento*
- Elle marchait au milieu des nuages = *Caminhava no meio das nuvens*
- La petite main soulevait le globe étoilé du monde = *A mãozinha erguia o globo estrelado do mundo*
- Un trou, ce village = *Esta aldeia é uma cafurna*

Quelques exemples de métaphores dans le domaine juridique pertinents pour notre étude :

| Métaphore juridique en français | Domaine source | Équivalent en portugais | Domaine source |
|---|------------------|--|------------------------|
| chambre criminelle | maison | <i>câmara penal</i> | maison |
| chambre préliminaire | maison | <i>juízo de instrução</i> | absence de sens figure |
| la grosse | spatial | <i>certidão relativa a um acto ou sentença</i> | absence de sens figuré |
| assiette uniforme | ustensile | <i>matéria colectável uniforme</i> | absence de sens figure |
| exploit d'assignation | antropomorphisme | <i>Notificação</i> | absence de sens figure |
| fruit civil | botanique | <i>fruto civil</i> | botanique |
| fiction juridique comme couverture | imaginaire | <i>ficção jurídica de cobertura</i> | imaginaire |
| Europe à géométrie variable | géométrie | <i>Europa de geometria variável</i> | géométrie |
| Europe à la carte | gastronomie | <i>Europa « à la carte »</i> | gastronomie |
| Europe à plusieurs vitesses | cinétique | <i>europa a várias velocidades</i> | cinétique |
| passerelle communautaire | architecture | <i>passarela comunitária</i> | architecture |

Dans cette phase de notre analyse encore au stade embryonnaire, la langue juridique, une langue très technique, emprunte en effet à de très nombreux registres de la langue courante et a ainsi largement recours au domaine architectural, anthropomorphique, géométrique, cinétique, spatial, ontologique, botanique, de la maison... Vivante, la langue juridique l'est assurément, à l'image des expressions suivantes : « personnalité juridique », « aliénation des biens », « contrat lésionnaire », autant d'expressions qui permettent une réelle personnification du droit. Il est enfin à noter dans notre corpus que la langue juridique a également souvent recours aux néologismes qui très souvent n'attendent pas la création d'un équivalent portugais, mais quand un équivalent est formulé, il est toujours intéressant de suivre en parallèle son évolution : *comitologie*, *gymnité* (*clause de ~*), *préclusion*, *forclusion*, *flexicurité* [*flexigurança*].

2.4 L'apparente aporie du changement linguistique

Nous l'avons compris, la langue de spécialité, elle aussi, est primitivement et fondamentalement métaphorique. Les concepts ne sont que d'anciennes métaphores usées, éteintes voire mortes. Stabilisés, ces libres regroupements deviennent des catégories qui peuvent prendre la forme de collocations, locutions, de cooccurrences, d'expressions idiomatiques, de catachrèses.

| | |
|--|--|
| <p>TERME: FRUITS CIVILS</p> | <p>INTERSECTION/ LIEN ANALOGIQUE</p> |
| <p>Définition: Les fruits civils qui sont obtenus grâce à un contrat dont le capital est l'objet, tels les loyers et autres revenus en argent procurés par une chose. 2. Par extension, qualification donnée à des biens qui constituent des produits au sens exact (dans la mesure où ils entament la substance de la chose qui les fournit), mais dont la production comme celle des fruits est périodique et régulière.</p> | <p>PRODUCTION, RENDEMENT.</p> |
| <p>TERME: LA GROSSE</p> | <p>INTERSECTION/ LIEN ANALOGIQUE</p> |
| <p>Déf. La « grosse » est le nom donné à la copie d'une décision de justice ou d'un acte notarié comportant la formule <u>exécutoire</u>. Elle est apposée par le Greffier de la juridiction qui a rendu la décision ou par le notaire qui a dressé l'acte contenant une reconnaissance de dette.</p> | <p>Le nom de « grosse » proviendrait de ce qu'à l'époque où les documents de justice étaient rédigés avec une plume d'oie, les commis des Greffiers et des Notaires étaient payés au <u>rôle</u>, de sorte que leur rémunération était d'autant plus élevée que la copie était longue. Ils avaient donc tout intérêt à écrire en grosses lettres. On parlait alors d'« écritures grossoyées ».</p> |
| <p>TERME: HUIS CLOS</p> | <p>INTERSECTION/ LIEN ANALOGIQUE</p> |
| <p>Déf.: L'expression s'utilise en droit pénal pour signifier que le public est exclu de la salle d'audience où se déroulent les débats. En procédure civile on emploie l'expression « <u>En Chambre du Conseil</u> » et on dira qu'en matière gracieuse les débats ont lieu en chambre du Conseil.</p> | <p>« Huis » vient de l'ancien français signifiant « porte » d'où le vocable « huisserie » pour désigner l'entourage des portes et des fenêtres.</p> |

L'évolution de la langue juridique dotée d'une personnalité à part entière résulte de l'ensemble des influences contingentes qui ont façonné son visage actuel. Comme tout organisme vivant, la langue est le produit d'une histoire avec ses hasards et ses ajustements. Dans le domaine juridique, une étude en diachronie s'impose pour pouvoir examiner avec plus de fiabilité le cheminement des termes, l'interaction entre les changements linguistiques et les milieux socioculturels. À la lumière de cette analyse, il apparaît clairement que l'étymologie est une forme d'étude diachronique. Pour comprendre le mécanisme du raisonnement analogique et son rapport avec la similitude, on doit s'intéresser impérativement à l'évolution du terme juridique où l'on recherche des situations potentiellement appariables avec la situation à analyser. Ainsi, l'objectif dans le raisonnement par analogie est alors d'exploiter un appariement adéquat effectué entre les éléments constitutifs des domaines source et cible, de manière à transférer un certain nombre de connaissances du premier vers le deuxième.

Conclusion et perspective

Nous avons présenté deux types d'approches effectuées en parallèle : une approche computationnelle pour la détection automatique des métaphores et pour l'aide à leur analyse, une approche linguistique comparative utilisant les mêmes corpus que ceux de l'étude computationnelle. Nous projetons maintenant une confrontation des approches afin d'asseoir la validité des modèles proposés et dans le but de créer un outil qui sera mis à disposition de la communauté scientifique.

Bibliographie

- Ji (H.), Ploux (S.), Wehrli (E.), 2003 : « Lexical knowledge representation with contonyms », dans *Proceedings of the 9th Machine Translation Summit*, p. 194 – 201.
- Ji (H.), Lemaire (B.), Choo (H.), Ploux (S.), 2008 : « Testing the Cognitive Relevance of a Geometric Model on a Word-Association Task: A Comparison of Humans, ACOM, and LSA », dans *Behavior Research Methods*, 40 (4), p. 926-934.
- Ben Maaouia (R.), *Vers un modèle pour la représentation de la sémantique des expressions dans le paradigme géométrique des Atlas sémantiques*, Rapport de mémoire de Master 2, Université Lyon 2, 2009. Oliveira (I.), Ploux (S.), 2008 : « Pour un traitement automatique de la métaphore », dans *Journée du CRTT: la métaphore en langues de spécialité*.
- Oliveira (I.), 2009 : *Nature et fonctions de la métaphore en science. L'exemple de la cardiologie*, L'Harmattan, Paris.

Patrick Drouin,
Frédéric Doll,
Observatoire de linguistique Sens-Texte,
Université de Montréal,
Montréal,
Canada,
patrick.drouin@umontreal.ca,
frederic.doll@nstein.com.

Potentiel terminologique, quel sens prendre ?

Résumé

Dans cet article, nous regardons le potentiel terminologique à la lumière des travaux récents en terminologie qui ont mis de l'avant des critères de sélection de termes destinés à l'humain. Ces travaux s'intéressent à l'unité terminologique dans son fonctionnement à la fois dans le texte et dans le lexique spécialisé et le décrivent sous l'angle de la sémantique lexicale. Nous désirons vérifier s'il est envisageable pour les concepteurs de logiciels d'acquisition automatique de termes d'exploiter ces critères pour l'évaluation du potentiel terminologique. Des résultats sont présentés et discutés sous l'angle de l'utilité pour le terminologue des données extraites par les techniques informatiques.

Mots-clés: terminologie; terminologie computationnelle; potentiel terminologique; linguistique de corpus; traitement automatique de la langue.

Introduction

La prise en charge ou le recours à l'informatique dans une activité terminologique ne cesse de prendre de l'ampleur. Le repérage des termes, étape importante à la fois pour le terminologue et pour de nombreuses disciplines connexes, demeure problématique. En effet, la nature référentielle des termes ne se laisse pas facilement manipuler à l'aide des techniques informatiques existantes. Comme l'indique Gaussier (2001), il n'existe toujours pas de définition du terme exploitable dans le cadre des travaux en acquisition automatique de la langue.

Afin de permettre aux systèmes informatiques de distinguer les unités terminologiques des unités non terminologiques dans les listes de candidats termes (CT) identifiés par les systèmes d'acquisition automatique de termes, Kageura et Umino (1996) ont proposé d'utiliser les concepts de « figement » (*unithood*) et de « potentiel » (*termhood*) terminologique. Le premier critère fait référence à la stabilité des CT recensés sur le plan de la syntaxe alors que le second se veut représentatif du potentiel d'une unité lexicale stable à se comporter comme un terme d'un domaine

particulier. Le figement terminologique peut être observé à l'aide de systèmes d'extraction automatique de termes basés sur des patrons morphosyntaxiques précis qui permettent de faire le recensement des unités lexicales dont les fréquences sont les plus élevées. Cependant, la prise en charge du potentiel terminologique est plus difficile à faire et les chercheurs en terminologie computationnelle ont exploité divers indices pour y parvenir, dont l'analyse du contexte immédiat (Frantzi et Ananiadou 1999), le recensement des unités simples qui constituent une unité lexicale complexe (Nakagawa et Mori 1998), l'exploitation de ressources terminologiques (Maynard et Ananiadou 2001), la parenté lexicale entre unités complexes (Drouin et Ladouceur 1994; Assadi et Bourigault 1996; Frantzi et Ananiadou 1997), la fréquence (Daille *et al.* 1994), l'importance d'un candidat terme pour un corpus (Ahmad *et al.* 1994; Kit 2002; Drouin 2003; Chung 2003; Gillam *et al.* 2005; Lemay *et al.* 2005), etc.

Dans cet article, nous regarderons le potentiel terminologique à la lumière des travaux récents en terminologie qui ont mis de l'avant des critères de sélection de termes destinés à l'humain. Ces travaux s'intéressent à l'unité terminologique dans son fonctionnement à la fois dans le texte et dans le lexique spécialisé et le décrivent sous l'angle de la sémantique lexicale. Nous désirons vérifier s'il est envisageable pour les concepteurs de logiciels d'acquisition automatique de termes d'exploiter ces critères pour l'évaluation du potentiel terminologique.

1 La sélection des unités terminologiques

Les manuels de terminologie présentent au terminologue divers indices pouvant le guider dans la sélection des unités lexicales représentatives d'un référent unique : la fréquence des unités lexicales, la nature principalement nominale des unités et la prévisibilité des patrons morphosyntaxiques. Ces indices sont directement pris en charge par les logiciels utilisés pour extraire les termes des textes spécialisés. Bien qu'utiles pour les humains, ces critères ne permettent pas de distinguer, par exemple, les syntagmes nominaux à caractère non terminologique (*téléphone bleu*) des unités ayant leur place dans un ouvrage terminologique (*téléphone portable, téléphone rouge*).

Un ouvrage récent publié par L'Homme (2004) a pour but de laisser de côté l'aspect plus flou de l'utilisation d'*indices* pour la sélection des termes par la proposition de *critères* tangibles. Au nombre de quatre, ces critères peuvent être énoncés de la façon suivante :

- L'unité lexicale extraite a un sens lié au domaine étudié;
- Pour les unités prédicatives, elles sont sélectionnées si les actants renvoient à des entités sélectionnées à partir du critère 1. Cependant, si l'unité peut se combiner avec des actants non spécialisés elle est écartée.
- Les dérivés morphologiques sémantiquement apparentés à des termes retenus en fonction des critères 1 ou 2 sont retenus.

- Les unités entrant dans une relation paradigmatique avec des termes sélectionnés en fonction des critères 1, 2 ou 3 sont conservées.

Ces critères simples permettent aux terminologues de systématiser leur décision et de justifier le choix des unités terminologiques sur des bases concrètes. Ils offrent aussi l'avantage de considérer le terme comme un élément qui évolue à la fois au sein du lexique et du texte et qui n'est pas isolé au milieu de structures conceptuelles difficiles à exploiter d'un point de vue terminologique. Ces critères sont des raccourcis pour la prise de décision. Ils peuvent éviter d'avoir à ouvrir une fois de plus un ouvrage de référence.

Les sections suivantes s'intéressent aux critères présentés précédemment et présentent notre vision de l'exploitation des propositions de L'Homme (2004) dans le cadre du traitement automatique de la langue.

1.1 Appartenance à un domaine de spécialité

L'appartenance à un domaine de spécialité d'une entité est une caractéristique difficile à évaluer sur le plan automatique puisque seuls le recours à la sémantique ou l'injection de connaissance humaine permettent de le déterminer. Cependant, l'utilisation de corpus spécialisés, élaborés pour des besoins précis de descriptions terminologiques et strictement délimités au préalable pour une sphère d'activité précise, nous permet d'envisager l'exploitation du corpus lui-même pour déterminer le statut d'une unité.

Comme nous l'avons mentionné auparavant, des travaux antérieurs sur le potentiel terminologique ont permis de dégager diverses techniques qui rendent possible l'identification du vocabulaire caractéristique d'un corpus. Si ce corpus couvre un seul domaine bien cerné, nous pouvons miser sur le fait qu'une partie du lexique de ce corpus sert à nommer les entités propres à ce domaine.

1.2 Nature des actants sémantiques

Le dépistage des actants sémantiques des unités prédicatives demeure lui aussi relativement difficile. Cependant, des heuristiques permettent de dégager des unités occupant un rôle syntaxique caractéristique des actants sémantiques potentiels. Par exemple, le dépistage des unités verbales décrivant une relation particulière/singulière du domaine d'étude pouvant être utilisées pour déterminer des unités terminologiques placées en position sujet ou objet :... *l'utilisateur navigue sur le Web à la recherche d'information...* Dans l'exemple, les actants sémantiques en position sujet et objet sont en petites majuscules alors que le prédicat sémantique est en caractère gras. Certains systèmes exploitent la structure syntaxique pour identifier les actants, d'autres exploitent la proximité des groupes nominaux autour d'un noyau verbal.

1.3 Parenté morphologique

Dans le cas des unités simples, une analyse morphologique peut permettre de dresser des listes de termes apparentés. Il faut cependant souligner que le deuxième volet de ce critère comporte une dimension sémantique que le traitement automatique de la langue ne peut valider. En effet, afin d'être retenues valides, deux formes morphologiquement reliées doivent aussi être apparentées sémantiquement. Le partage d'une chaîne de caractères commune permet d'inférer que les termes *programmation*, *programmeur*, *reprogrammer* sont probablement reliés sémantiquement les uns aux autres.

Dans le cas des unités complexes, les chercheurs (Drouin et Ladouceur 1994; Assadi et Bourigault 1996; Frantzi et Ananiadou 1997) se sont déjà intéressés à la réutilisation du matériau lexical pour l'évaluation du potentiel terminologique. La structure binaire (*tête* et *expansion*) des termes complexes ([*scie à cadre*] [*à lame unique*]) et leur division binaire récursive ([*scie [à cadre]*] [*à lame [unique]*]) rend possible la construction d'ensembles de termes apparentés par le partage d'une tête ou d'une expansion. Ainsi, on peut considérer qu'à l'intérieur d'un domaine de spécialité les termes qui partagent des éléments lexicaux communs sont sémantiquement apparentés (*scie à cadre à lame unique*, *scie à cadre pour bois de mine*, etc.)

1.4 Liens paradigmatiques

La technique la plus simple pour l'exploitation de lien paradigmatique réside dans l'exploitation de ressources terminologiques (Frantzi et Ananiadou 1999) pour l'établissement de liens sémantiques sur la base de la forme. Une ressource généraliste comme Wordnet pourrait aussi fournir de l'information de base sur le lien qui unit deux formes.

En plus de sources d'informations externes au corpus, le dépistage de liens de synonymie et d'antonymie peut être conduit grâce à une technique qui consiste à reprendre l'analyse distributionnelle à la Harris en exploitant les termes en contexte et la statistique pour quantifier ce phénomène (Bourigault 2002).

Une telle opération peut s'effectuer directement sur la liste des résultats pour les séries dont les expansions sont fixes avec des têtes variables ou à l'inverse des termes dont seule l'expansion varie. Dans certains cas, les têtes seront utilisées avec des formes qui sont des synonymes ou des antonymes (*clapet à obturateur à déplacement angulaire*, *clapet à obturateur à déplacement linéaire*; *synonymie parfaite*, *antonymie parfaite*).

Pour les unités terminologiques complexes de nature compositionnelle, la nature du lien sémantique entre la tête et l'expansion peut varier énormément: *crayon à souder* (fonction), *tête de moteur* (partie-tout), *imprimante à laser* (mode de fonctionnement). De nombreuses études sur le sujet ont été réalisées dans le cadre de recherches sur la nature des unités nominales complexes en terminologie

(Lotte 1981 ; Sager 1990 ; L'Homme 1996) et en traduction automatique (Fontenelle 1997 ; Ananiadou et Mc Naught 1990). À notre connaissance, l'application de ces recherches à la terminologie computationnelle, plus particulièrement à l'évaluation du potentiel terminologique, bien que prometteuse dans des cas très ciblés, se fait attendre.

2 Méthodologie

Afin de mettre en place concrètement les quatre critères cités précédemment, nous avons développé des solutions pour « coller » au mieux à ces exigences et nous permettre de dégager des avenues de recherche en prenant en compte l'utilisabilité d'un outil encore imparfait.

2.1 Prise en charge du lien avec le domaine

Si le terminologue peut aisément se faire une idée du potentiel terminologique d'un terme en le reliant au domaine étudié, la tâche est plus ardue pour la machine. Le contour d'un domaine est généralement flou et le réduire à une liste finie de concepts désignés par des termes ou de relations est une opération futile et vouée à l'échec étant donné que les domaines sont des mondes ayant des intersections nombreuses. Cependant, une définition négative de l'appartenance à un domaine est envisageable : « fait partie intégrante d'un domaine de spécialité, un terme ou une relation absent ou faiblement représenté dans la somme des domaines existants ».

En compilant une base de données de termes tirés d'un corpus couvrant une variété de domaines, il est possible de pondérer l'importance d'un terme pour un domaine en comparant sa fréquence relative dans le corpus d'analyse et celle de la base de données générale. Le logiciel TermoStat, qui nous a sert à explorer certaines pistes de recherche, fonctionne sur ce principe.

Un module extérieur au logiciel attribue à chaque mot (nom, verbe, adjectif...) une partie du discours. Un filtre basé sur une série d'expressions régulières est appliqué sur chacune des phrases du corpus afin d'en extraire les candidats termes (CT). Les matrices syntaxiques prédéfinies les plus courantes sont :

- Nom : *définition, dictionnaire*
- Nom + adj : *champs sémantique, définition lexicale*
- Nom + prép + nom : *partie du discours, dictionnaire de langue*
- Nom + prép + nom + adj : *complément d'objet direct, principe de compositionnalité sémantique*
- Nom + part pass : *variation liée, langue écrite*
- Nom + adj + prép + nom : *structuration sémantique du lexique, approche sémiotique du langage*

L'étiquetage des termes dans un corpus volumineux couvrant des domaines multiples est préalablement lancé et les fréquences de chacun des termes trouvés sont compilées. La même opération d'étiquetage est ensuite lancée sur le corpus d'analyse. La mise en parallèle du corpus d'analyse et du corpus général (corpus de référence) est un moyen pour déterminer sa représentativité dans un corpus général et inférer, lorsque cette représentativité est faible ou nulle, l'appartenance à un domaine de spécialité, au domaine étudié.

Pour évaluer cette appartenance à un domaine de spécialité, les fréquences dans les deux corpus sont comparées statistiquement à l'aide du rapport de chance (*log odds ratio*). L'utilisation de cet indice statistique pour pondérer chacun des candidats permet de prendre en compte la fréquence relative des concepts dans chacun des corpus. Des tests comparatifs ont été effectués et les tests effectués avec le *log odds* donnaient de meilleurs résultats (Drouin et Doll 2008).

2.2 Exploitation de la cooccurrence

Dans les domaines spécialisés, les unités prédicatives sélectionnent dans un bassin relativement restreint leurs actants sémantiques. Ce processus de restriction s'exprime sur le plan syntaxique par une certaine forme de redondance, plus précisément, de cooccurrence lexicale. Pour déterminer les actants d'une unité prédicative, le logiciel TermoStat, à défaut de procéder à une analyse syntaxique fiable sur l'ensemble de la phrase, calcule le degré de cohésion entre deux termes apparaissant dans une relation de proximité dans la phrase (proximité nom/verbe = relation sujet ou objet). Deux listes de paires de termes (nom/verbe) et (verbe/nom) sont extraites et compilées avec leur fréquence grâce à un algorithme simple de recherche du nom le plus proche à gauche et à droite d'un verbe.

Pour chacune des paires de termes extraites, le logiciel applique un indice de vraisemblance (*log-likelihood*) qui attribue, à partir de a) la fréquence de l'association, b) la fréquence des associations avec le premier terme seulement, c) la fréquence des associations avec le second terme seulement et d) la fréquence des autres paires, un poids à la paire de termes. Ce calcul permet d'écarter ainsi les actants non spécialisés (qui se retrouvent fréquemment dans d'autres associations)

2.3 Structuration terminologique

Nous envisageons la prise en charge des liens paradigmatiques entre termes à l'aide de deux techniques : la détection de phénomènes liés à la dérivation morphologique et la structuration terminologique.

La détection de la dérivation morphologique des termes simples, bien qu'envisageable automatiquement avec une précision satisfaisante, n'est pas effectuée par le logiciel. Elle resterait cependant limitée et contrainte. Pour le moment, le logiciel utilisé contourne cet obstacle en rendant à l'utilisateur la supervision de la dérivation existante pour chacun des termes. En triant les listes de candidats par ordre alphabétique sur le premier terme, des groupes de dérivations morphologiques

qui peuvent servir à déterminer le potentiel terminologique émergent en bloc. D'autres pistes exploratoires sont envisageables afin de rapprocher physiquement (dans une même page, sous forme de nuage, de liste, de tri sur la fin des mots...) des termes issus d'un morphème commun. Des calculs sur la proximité graphique d'un ensemble de termes sélectionnés pourraient donner des résultats exploitables.

Un processus de décomposition des termes complexes recensés par le logiciel a pour but de construire un réseau de parenté entre les termes navigable par l'utilisateur. Ce réseau facilite la visualisation du partage d'éléments lexicaux (tête ou expansion) dans la liste des unités extraites par le logiciel.

Chaque groupe nominal qui se compose d'un nom et d'un adjectif est stocké à la fois dans une liste de noms ayant une relation avec un adjectif et dans une liste d'adjectifs ayant une relation avec un nom. Ce double stockage des groupes nominaux comme « milieu familial », « milieux défavorisés » et « enfant défavorisé » permet de naviguer du premier au troisième grâce au partage du nom en commun des deux premiers et de l'adjectif en commun des deux derniers.

2.4 Relation paradigmaticque des unités

C'est sous la forme de graphe que les relations paradigmaticques sont mises en évidence par le logiciel TermoStat. Après avoir effectué les étapes d'extraction de patrons syntaxiques, de constitution de paires de termes cooccurrents ou syntaxiquement liés et de pondération de ces paires de termes, il a été possible de relier chacun des constituants de ces paires en un graphe.

Les paires [Nom₁-Adjectif₁] et [Nom₂-Adjectif₁] ayant reçu un score de cohésion suffisant deviennent des composants du graphe. Le lien, inexistant dans le texte entre Nom₁ et Nom₂, devient visible par l'intermédiaire de l'adjectif qu'ils partagent.

Si Nom₁ et Nom₂ partagent un deuxième adjectif en commun (Adjectif₂), le paradigme dans lequel ils se retrouvent est détectable et infère une relation entre Nom₁ et Nom₂ qui peut parfois être difficilement étiquetable, mais indéniablement vecteur de sens et garant d'un potentiel terminologique accru pour ces unités qui constituent ces micrographes.

3 Résultats

La présente section illustre les résultats de la mise en œuvre des diverses techniques décrites dans la section 4 afin de mettre en lumière le potentiel terminologique des candidats termes sur la base des critères énoncés dans la section 3.

3.1 Appartenance à un domaine spécialisé

Afin d'obtenir des exemples variés, nous avons procédé à l'extraction des candidats termes dans trois corpus traitant de sujets différents : la lexicologie et la sémantique

lexicale (LEXICO), le terrorisme et les engins explosifs improvisés (TERRO) et la mécanique automobile (AUTO).

| LEXICO (FR) | TERRO (EN) | AUTO (EN) |
|-----------------------|-----------------------------|---------------------------|
| lexie | <i>ieds</i> | <i>axle</i> |
| signe linguistique | <i>jammer</i> | <i>rod</i> |
| fonction lexicale | <i>x-ray</i> | <i>carburetor</i> |
| mots-forme | <i>electronique jammer</i> | <i>tire</i> |
| prédicat | <i>countermeasure</i> | <i>vapor</i> |
| collocation | <i>firing train</i> | <i>air-fuel</i> |
| lexème | <i>x-ray method</i> | <i>aluminum</i> |
| actant | <i>ied reconnaissance</i> | <i>throttle vave</i> |
| sens lexical | <i>mine</i> | <i>modern vehicle</i> |
| affixe | <i>electromagnetic wave</i> | <i>flywheel</i> |
| relation sémantique | <i>military explosive</i> | <i>impeller</i> |
| sens linguistique | <i>detection technique</i> | <i>drive shaft</i> |
| définition analytique | <i>gas chromatography</i> | <i>combustion chamber</i> |
| champ sémantique | <i>enemy force</i> | <i>output shaft</i> |
| genre prochain | <i>secure element</i> | <i>rear axle</i> |
| unité lexicale | <i>ied training</i> | <i>pinion</i> |

Tableau 1

La première colonne présente des résultats obtenus pour un corpus de langue française alors que les deux dernières sont représentatives de données extraites en anglais. Dans la majorité des cas, on peut rapidement évaluer la pertinence des unités recensées pour les domaines en question. Ces listes sont cependant triées en ordre décroissant et les premières unités (celles qui sont présentées ici) sont souvent valides alors que celles situées en fin de liste sont moins pertinentes. Le problème majeur consiste donc à trouver un moyen d'établir un seuil ayant pour but d'élaguer de la liste les unités qui ne sont pas reliées au domaine.

3.2 Nature des actants sémantiques

Le tableau 2 présente les résultats obtenus pour deux verbes tirés du corpus LEXICO. Les termes suivis de *-S* ont été observés en position sujet et ceux suivis de *-O* en position objet. Dans ce contexte, nous cherchons à établir le potentiel terminologique des verbes.

| | |
|-----------------|----------------|
| véhiculer | contrôler |
| forme-S | base-S |
| idée-O | combinatoire-O |
| signification-O | combinaison-O |
| nombre-O | collocation-O |
| sens-O | construction-O |
| présupposé-O | |

Tableau 2

Dans les deux cas, on peut établir l'intérêt terminologique des verbes de façon heuristique à partir de la présence des noms dans les listes de la section précédente (liste de termes du domaine). Par contre, afin d'établir de façon catégorique que ces verbes sont des unités terminologiques, il faut pousser l'analyse plus loin et s'assurer que le sens véhiculé par ces verbes est différent de celui qu'ils adoptent en dehors du domaine de la lexicologie. Une approche logicielle ne permet pas d'effectuer cette vérification, mais peut au moins présenter de manière concise et limitée certains cas à l'utilisateur terminologue qui est en mesure de trancher.

3.3 Parenté morphologique

L'exploitation de la parenté morphologique sur la base du partage d'une chaîne de caractères commune ou d'un morphème commun permet de construire assez rapidement des familles de termes : *lexie*, *lexical*, *lexicalisation*, *lexicologie*, *lexicologue*, *lexicologique*, *lexicographe*, *lexicographie*, *lexicographique*, *lexème*, *lexicométrie*, *lexie*, *lexique*. La présence au sein d'un corpus centré sur un domaine d'un bon nombre de ces termes apparentés morphologiquement est probablement un bon indicateur du potentiel terminologique des unités qui constituent la famille.

En ce qui a trait aux unités complexes, la mise en correspondance des unités ne se fait pas sur la base du partage d'une chaîne commune, mais plutôt sur le partage d'un ou plusieurs éléments lexicaux. Le tableau 3 illustre ce phénomène.

| | |
|--------------|------------|
| sens | lexical |
| lexical | fonction |
| linguistique | sens |
| logique | unité |
| notion | sémantique |
| grammatical | signifiant |
| implicite | signe |
| élément | réseau |
| ... | ... |

Tableau 3

La décomposition des divers termes complexes faisant appel à *sens* (*sens lexical*, *sens linguistique*, *notion de sens*, *sens implicite*, etc.) permet de relier ce terme à un ensemble d'autres termes en considérant que des éléments lexicaux sont communs à deux termes. Ainsi, de *sens lexical*, on peut naviguer aux termes *fonction lexicale*, *unité lexicale*, *sémantique lexicale*, *réseau lexical*, etc.

La mise en commun récursive de ces relations entre termes sur une base purement lexicale au sein d'un graphe facilite la visualisation de cette parenté. Plus encore, le graphe peut également être « centré » (par une intervention manuelle) sur une branche périphérique pour montrer un réseau ayant pour centre la nouvelle sélection. La figure 1 illustre le réseau terminologique élaboré autour de la tête *réseau* sur le corpus LEXICO.

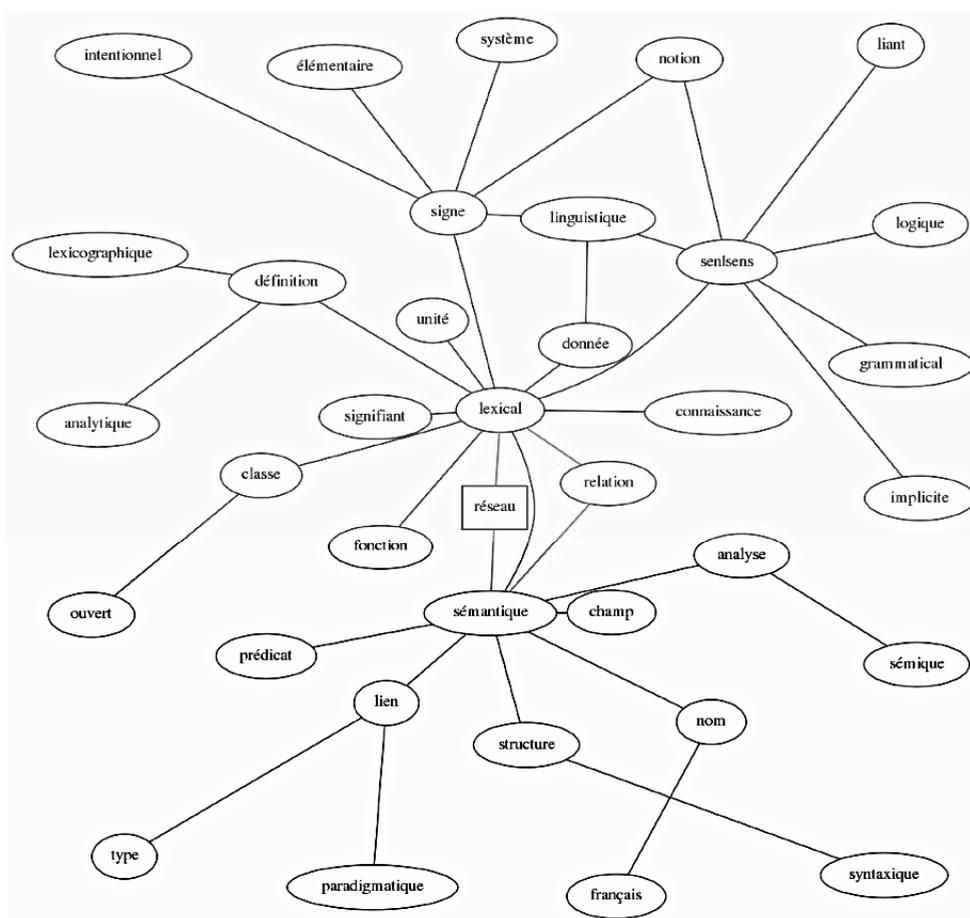


Figure 1

Nous croyons que la consultation de ces graphes par le terminologue peut l'aider dans l'évaluation du potentiel terminologique des candidats recensés par le logiciel. Dans un même ordre d'idée, la structure de ces graphes pourrait conduire le

logiciel à détecter certains phénomènes comme l'apparition de cliques (circuits fermés) au sein des graphes comme dans le cas de *réseau-lexical-relation-sémantique* ou encore *signe-notion-sens-lexical-linguistique*, etc.

3.4 Liens paradigmatiques

Le passage à une représentation sous forme de graphe présenté dans la section précédente permet aussi de mettre en évidence des liens de nature paradigmatique. L'analyse issue de la cooccurrence autour du verbe *dénoter* dans le corpus LEXICO donne lieu au graphe qui suit (figure 2).

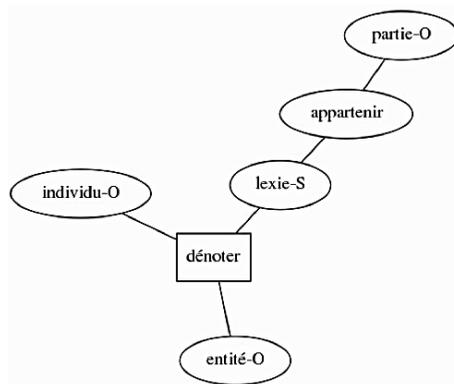


Figure 2

On note ici que *dénoter* sert de pivot et nous permet de relier *individu* et *entité* qui correspondent aux actants sémantiques dans une structure du genre *X dénote Y*. L'instanciation d'*entité* et d'*individu* dans le corpus pourrait conduire à l'identification de séries paradigmatiques comme *entité-objet-chose* ou *individu-personne-être*.

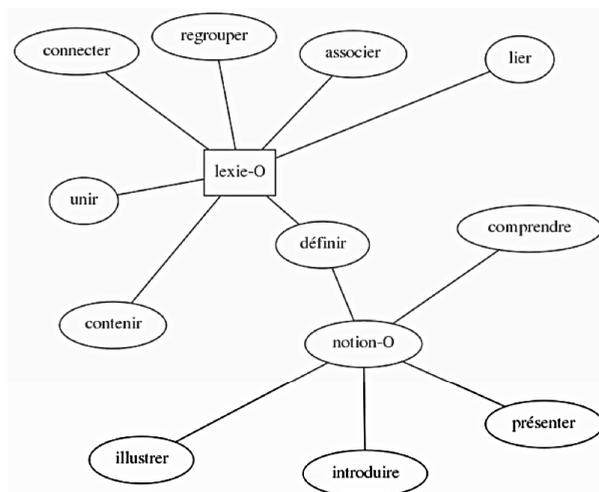


Figure 3

De la même façon, le graphe précédent (figure 3) met en évidence une parenté sémantique évidente entre les verbes *illustrer*, *présenter*, *introduire* (*porter quelque chose à l'attention, à la connaissance de quelqu'un*).

La construction de graphes à partir des résultats de la structuration terminologique va conduire à la description de relations sémantiques très variées. Le graphe suivant (figure 4), construit autour du terme *transmission* dans le corpus AUTO laisse voir une mise en opposition de deux types de transmissions (co-hyponymie), *les transmissions automatiques* et *les transmissions manuelles*.

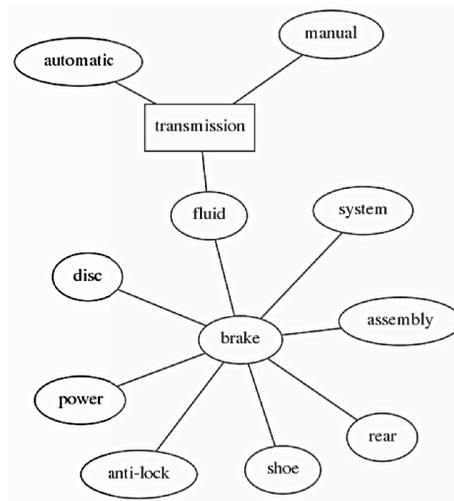


Figure 4

Le même graphe comporte une relation d'holonymie entre le frein (*brake*) et ses composantes (*disc*, *shoe*) et diverses relations sémantiques (*brake – rear*, *brake – fluid*, etc.) qui sont cependant toutes illustrées de la même façon. Encore une fois, l'utilisateur du logiciel peut faire la part des choses et assez rapidement constater l'intérêt terminologique de ces relations paradigmatiques. Étant donnée la diversité des relations dépistées, il demeure par contre très difficile d'envisager l'inclusion de tel indice dans une pondération visant à évaluer de manière automatique le potentiel terminologique.

Conclusion

Dans cet article, nous nous sommes arrêtés au concept de potentiel terminologique à la lumière des travaux récents en terminologie. Ces derniers ont mis de l'avant des critères de sélection de termes destinés à l'humain. Nous avons tenté de vérifier s'il est envisageable pour les concepteurs de logiciels d'acquisition automatique de termes d'exploiter ces critères pour l'évaluation du potentiel terminologique. Du point de vue du traitement automatique de la langue, les outils et les techniques que nous avons utilisés ne sont pas nouveaux. Cependant, ces pistes n'ont jamais

été exploitées pour l'évaluation du potentiel terminologique des candidats termes identifiés par un outil automatique. Les résultats que nous avons obtenus pour chacun des critères sont encourageants.

L'évaluation du potentiel terminologique est une tâche de terminologue. C'est l'homme qui appose son sceau sur l'entrée ou non d'une unité terminologique dans une base de données. Les mécanismes qui poussent vers l'acceptation ou le rejet d'une unité sont très complexes et c'est autant la connaissance du domaine que l'expérience sociale du professionnel qui joue un rôle dans cette sélection. Ces mécanismes subtils conduisent cependant à un verdict binaire *oui – non*.

Le défi que nous nous sommes fixé n'était pas de se substituer au décisionnaire, mais, d'une part, de vérifier si des logiciels d'acquisition automatique de termes pouvaient augmenter la perception du domaine d'étude en proposant des paradigmes estompés par l'éloignement (les dérivés morphologiques) et, d'autre part, d'alimenter l'expérience sociale en utilisant la machine comme un lecteur compulsif de grandes quantités de textes qui serait ensuite en mesure de dégager des domaines et de pondérer chaque unité rencontrée.

Nous croyons que ces avenues sont envisageables, que la nuisance trop bien connue de l'aide informatique peut être évitée en ne proposant que des pistes bien défrichées qui doivent cependant pouvoir être à remises en question. Le défi demeure de remettre tout ça dans une mesure du potentiel terminologique. Est-ce la bonne avenue ? Répondre à cette question demeure difficile. Cependant, nous sommes d'avis que nous ne devons pas uniquement chercher à quantifier le potentiel terminologique à l'aide de méthodes statistiques qui mesurent uniquement le rappel et la précision, mais nous orienter vers un autre paradigme qui est celui de l'intégration des indices pris en charge par l'humain et de l'utilité pour le terminologue des données proposées.

Bibliographie

- Ahmad (K.), Davies (A.), Fulford (H.) et Rogers (M.), 1994 : « What's in a Term ? The Semi-automatic Extraction of Terms from Text », dans Pöschhacker (F.) et Kaindl (K.), (éd.), *Translation Studies*. Amsterdam : Benjamins, 1994.
- Ananiadou (S.) et McNaught (J.), 1990 : « Treatment of compounds in a transfer-based MT system », dans *Proceedings of 3rd international conference on theoretical and methodological issues in machine translation*, Austin, p. 57-63.
- Assadi (H.) et Bourigault (D.), 1996 : « Acquisition et modélisation des connaissances à partir de textes : outils informatiques et éléments méthodologiques », dans *Actes du 10e congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA'96)*, Rennes, p. 505-514.
- Bourigault (D.), 2002 : « Upery : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus », dans *Actes de la 9e conférence annuelle sur le Traitement Automatique des Langues (TALN 2002)*, Nancy, p. 75-84.
- Chung (T. M.), 2003 : « A corpus comparison approach for terminology extraction », dans *Terminology*, vol. 9, n° 2, p. 221-246.

- Daille (B.), 1994 : « Extraction de noms composés terminologiques du domaine des Télécommunications », dans *5es Journées ERLA-GLAT (Études et Recherches Lexicales Appliquées)*, Brest, 13 p.
- Drouin (P.), 2003 : « Term extraction using non-technical corpora as a point of leverage », dans *Terminology*, vol. 9, n° 1, p. 99-117.
- Drouin (P.) et Doll (E.), 2008 : « Quantifying Termhood Through Corpus Comparison », dans *Terminology and Knowledge Engineering (TKE-2008)*, Copenhagen Business School, Copenhagen, p. 191-206.
- Drouin (P.) et Ladouceur (J.), 1994 : « L'identification automatique de descripteurs complexes dans des textes de spécialité », dans *Proceedings of the Workshop on Compound Nouns: Multilingual Aspects of Nominal Composition*, Genève, ISSCO, p. 8-28.
- Fontenelle (T.), 1997 : « Using a bilingual dictionary to create semantic networks », dans *International Journal of Lexicography*, vol. 10, n° 4, p. 275-303.
- Frantzi (K.) et Ananiadou (S.), 1999 : « The C-value & NC-value Domain Independent Method for Multi-Word Term Extraction », dans *Journal of Natural Language Processing*, vol. 6, n° 3, p. 145-179.
- Frantzi (K.T.) et Ananiadou (S.), 1997 : « Automatic Term Recognition Using Contextual Cues », dans *Proceedings of the 3rd DELOS Workshop*, Zurich, 8 p.
- Gaussier (É.), 2001 : « General considerations on bilingual terminology extraction », dans *Recent advances in Computational Terminology*, Bourigault *et al.* (éd.), Amsterdam/Philadelphia, John Benjamins Publishing Company, p. 167-183.
- Gillam (L.), Tariq (M.) et Ahmad (K.), 2005 : « Terminology and the Construction of Ontology », dans *Terminology*, vol. 11, n° 1, p. 55-81.
- Kageura (K.) et Umino (B.), 1996 : « Methods of Automatic Term Recognition: A review », dans *Terminology*, vol. 3, n° 2, p. 259-289.
- Kit (C.), 2002 : « Corpus Tools for Retrieving and Deriving Termhood Evidence », dans *5th East Asia Forum of Terminology*, p. 69-80.
- Lemay (C.), L'Homme (M.-C.) et Drouin (P.), 2005 : « Two Methods for Extracting « Specific » Single-word Terms from Specialized Corpora: Experimentation and Evaluation », dans *International Journal of Corpus Linguistics*, vol. 10, n° 2, p. 227-255.
- L'Homme (M.-C.), 2004 : *La terminologie: principes et techniques*, Montréal: Les Presses de l'Université de Montréal.
- L'Homme (M.-C.), 1996 : « Sélection des prépositions dans les termes complexes Nom (Prép.) Nom à partir de leur structure conceptuelle », dans *Cahiers de Lexicologie*, vol. 68, n° 1, p. 25-43.
- Lotte (L.), 1981 : « Principes d'établissement d'une terminologie scientifique et technique », dans *Fondements théoriques de la terminologie*, Siforov, Vladimir Ivanovich (éditeur), Groupe interdisciplinaire de recherche scientifique et appliquée en terminologie, Québec: GIRSTERM, p. 1-53.
- Maynard (D.) et Ananiadou (S.), 2001 : « Term extraction using a similaritybased approach », dans *Recent advances in Computational Terminology*, Bourigault *et al.* (éd.), Amsterdam/Philadelphia, John Benjamins Publishing Company, p. 261-278.
- Nkagawa (H.) et Mori (T.), 1998 : « Nested Collocation and Compound Noun for Term Extraction », dans *Computerm'98. First Workshop on Computational Terminology. Proceedings of the Workshop*, 15 août 1998, Université de Montréal, p. 64-70.
- Sager (J.-C.), 1990 : *A Practical Course in Terminology Processing*, Amsterdam/Philadelphie, John Benjamins Publishing Company.

Mosbah Saïd,
Université de Carthage, UR/00/0201,
Tunis,
Tunisie,
Said_m_tn@yahoo.fr.

Traduire les théories linguistiques : l'exemple de la psychosystématique du langage

Résumé

Nous nous proposons, ici, d'étudier les questions que soulève la traduction de la psychosystématique vers l'arabe. Cette théorie établissant un certain rapport entre la pensée et le langage, censé être universel, est élaborée dans le cadre d'un univers linguistique bien déterminé. Sa transposition dans un autre système dont l'origine, les structures et le fonctionnement sont différents, nécessite « une réécriture » (sur le plan des illustrations). Elle doit être accompagnée d'un effort de conceptualisation systématique à partir de la langue cible visant à ancrer les concepts dans des contextes spécifiques de cette langue.

Mots-clés : traduction ; métalangue ; théorie linguistique.

On peut définir une théorie linguistique comme un réseau de concepts forgés dans une langue, exprimés par une terminologie conçue dans le cadre de cette langue et appliquée à cette langue elle-même.

Traduire cette théorie dans une autre langue pose la problématique de l'adéquation terminologique vue sous l'angle des contraintes de conceptualisation première et sous l'angle de la différence de l'objet-langue qui a servi ou va servir de support à cette conceptualisation.

Sur le plan terminologique strict, cette contrainte provient du fait que l'équivalent peut respecter ou ne pas respecter le contenu notionnel du terme initial. Sur le plan conceptuel, il y a toujours un risque de projeter ces contenus dans leurs plus fins détails sur la langue cible.

Ce sont ces deux aspects que nous tenterons d'étudier en soulevant les problèmes que pose la traduction vers l'arabe d'une théorie connue par son haut degré d'abstraction (la psychosystématique).

1 Problématique de l'équivalence terminologique

Le terme métalinguistique peut se servir d'unités existantes dans la langue qui sont souvent polysémiques et les doter de contenus notionnels propres comme il peut être forgé, souvent par composition, à partir des éléments lexicaux disponibles.

Dans le premier cas, émergeant à partir d'une vision synthétique des faits observés ou décrits dans l'objet d'étude (langue, discours), le concept lié à ce terme sera considéré comme une nouvelle acception du mot (ou *terme* en langue de spécialité). De ce fait, il s'apparente plus à l'une des acceptions déjà instituées en langue avec laquelle il partage un ou plusieurs traits sémiques et par là même, dans certains cas, ce sont ces mêmes traits qui fondent les rapports d'antonymie ou de synonymie et du mot et du terme.

Soit par exemple, le terme *incidence* qui, chez Guillaume, renvoie à deux concepts différents selon qu'il couvre l'opposition « incidence/décadence » ou l'opposition « apport/support ». Du côté de la langue, le mot a au moins trois significations :

- 1. (vx) Ce qui arrive souvent, survient.
- 2. Rencontre d'une ligne, d'un corps et d'une autre ligne, d'une surface.
- 3. Conséquence, effet, influence; plus particulièrement « effet de la charge fiscale sur une personne ». (*Le Petit Robert*)

Dans l'ensemble, on peut ramener ces acceptions à trois noyaux sémiques à savoir :

- mouvement du futur vers le présent (ce qui arrive);
- effet;
- rencontre, point de rencontre.

Dans la première acception employée par Guillaume, la définition d'« incidence » est corrélée à celle de « décadence ». Les deux concepts renvoient à « chacun des deux moments déterminés par l'échéance de l'événement au temps d'univers: l'événement vu arrivant au temps se situe au niveau d'incidence (niveau α , procès en accomplissement), l'événement vu arrivé au temps se situe au niveau de décadence (niveau ω , procès accompli) » (Boone et Joly 2004 : 232) :

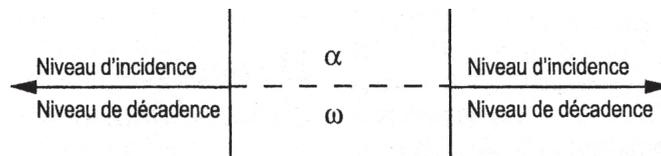


Figure 1

Ainsi, l'incidence et la décadence sont définies comme des intervalles d'espace-temps : l'un prenant départ dans le futur et s'arrêtant à la limite de la parcelle future du présent ; le second prenant départ au début de la parcelle passée du présent et allant vers le passé.

Il est clair que la notion d'incidence s'apparente au premier noyau sémique (vieux en langue) et qu'elle exclut du même coup les idées de « rencontre » et d'« effet » ; fait que justifie également son antonyme « décadence » dont le contenu notionnel repose sur l'inversion de ce même noyau sémique (ce qui arrive ≠ ce qui s'en va).

Dans sa recherche de l'équivalent terminologique, le traducteur doit donc d'abord déterminer le lien entre le contenu notionnel du terme et son substrat lexical en L1 avant de transposer ce même lien (avec tout son réseau relationnel) en L2. Dans le cas contraire et s'il s'en tient à ce qu'offrent les dictionnaires bilingues généraux, il risque de trouver son choix limité à des équivalents transposés à partir de l'idée d'« effet » et il traduira « incidence » par « *ʔinʕika:s* » (littéralement, « reflet », « réflexion ») ou à partir de l'idée de « rencontre » et il traduira le terme par *tala:qi:* (littéralement : « rencontre » avec idées de réciprocité ou de récurrence) ou encore par *ʔisqa:t* (« projection ») qui s'éloigne de la signification recherchée.

À notre sens, cette opposition entre « incidence » et « décadence » peut être traduite par le couple *alʔa:ti:* (*le venant) ou *alqa:dim* (*l'arrivant)/*almunqadi:* (ce qui passe).

Dans le second emploi, « incidence » est défini comme « un mécanisme qui régit la relation entre apport de signification et support de signification. Elle est la faculté qu'ont les mots de se référer à un support. » Guillaume distingue dans ce sens deux configurations possibles : l'incidence interne « qui ne sort pas de ce que le mot connote, c'est-à-dire lorsque l'apport et le support de signification ne peuvent être dissociés » (le cas du substantif) et l'incidence externe « lorsque l'apport et le support sont dissociés et qu'il y a, par conséquent, incidence à un support pris en dehors du mot (cas du verbe et de l'adjectif) » (*ibid.* : 234-235).

L'idée d'incidence peut ici être rapprochée en partie de la relation de réflexivité en mathématiques. Soit la relation « avoir comme support... » ; dans le cas de l'incidence interne (le substantif), il s'agit bien d'une relation de réflexivité au sens mathématique du terme :

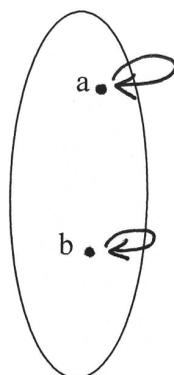


Figure 2

Dans le cas de l'incidence externe, la relation devient anti-symétrique :

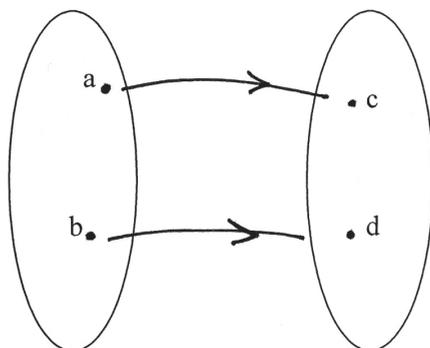


Figure 3

L'idée d'effet n'est donc pas loin dans ce type d'incidence. Selon les catégories de discours, chaque élément trouve son effet soit dans lui-même, soit à l'extérieur de lui-même. La traduction du terme dans cette acception par *ἠνσικα:s* nous paraît ici judicieuse (Mejri à paraître).

Mais, dans le second cas et à part cette sélection à partir du multiple lexical, la terminologie peut également générer le différent à partir du semblable en langue. C'est précisément le cas de composition de deux termes à partir d'éléments quasi synonymiques : il s'agit de *chronogenèse* et *chronogénie*. Si le premier élément est le même (*chrono-* = temps), la différence entre les deux suffixes *genèse* et *génie* n'est nullement évidente :

- *genèse* : naissance, formation, production ;
- *génie* : production, formation (*Le Petit Robert*).

La différence est donc à saisir au niveau des traits conceptuels des deux concepts formés.

La chronogenèse est définie comme « l'opération de pensée formatrice de l'image-temps » (Guillaume 1971 : 133) ; la chronogénie comme « la construction du langage dans le temps » (Guillaume 1964 : 33) et Guillaume y voit une similitude avec la « glossogénie » (construction du langage dans le temps).

Pour forger ses termes équivalents en L2, le traducteur doit d'abord attribuer une signification plus ou moins distinctive aux éléments *-genèse* et *-génie*. Cette signification, il va la chercher dans la différence entre les deux notions formées en L1. De prime abord, il s'agit dans les deux cas d'un processus de formation du temps. Mais, ce qui nous paraît distinctif, c'est que le premier processus est présenté comme ayant un agent explicite (*la pensée formatrice*) alors que le second, défini comme construction, ne l'a pas, du moins explicitement. En effet, si la saisie initiale de la chronogenèse se situe à une distance plus ou moins grande de son origine et relève du déjà construit par la pensée (la notion d'agent est donc présente dès le début pensable), le moment initial de la chronogénie (dit *causation observe*; *causation* dans le sens strict d'effectivité) relève de l'indicible. Autrement dit, « du dicible en construction **lui-même**... pure visibilité mentale » (Boone et Joly 2004 : 92), sans agent extérieur.

Cette distinction est cruciale pour le traducteur car, selon le cas, elle lui permettrait de choisir, parmi plusieurs vocables quasi synonymiques en L2, celui dont le schème exprime l'agentivité ou l'auto-agentivité. Ainsi, pour *chronogenèse*, il choisirait *tawli:d zamani*: (= litt. *engendrement*; (Mejri 2009 : à paraître)) ; *tawli:d* requiert sa forme du schème *tafci:l* qui sous-tend un agent externe potentiel. Pour *chronogénie*, il pourrait choisir *attakawwun-izzamani*: (= litt. *construction temporelle*) où *takawwun* épouse le schème *tafaʕul* rendant l'idée d'auto-agentivité.

Toutefois, tout cela n'a pas de sens si les illustrations en L2 ne rendent pas l'essence des concepts de L1. Ceci pose en plein toute la problématique de l'universel et du spécifique dans le cadre d'une théorie qui, selon les termes de Roch Valin (dans Guillaume 1984 : XIII), serait « capable non seulement d'embrasser tous les langages humains mais aussi de les situer les uns par rapport aux autres » et dont l'application a pour cadre une langue particulière.

2 Problèmes de l'application des concepts à la langue cible

Peu nombreux, les passages où Guillaume a évoqué la langue arabe (ou les langues sémitiques dont elle fait partie) relèvent de la linguistique générale. Il s'agit surtout de rappeler que, dans les langues sémitiques, ce sont les consonnes qui donnent au vocable sa substance-matière et que les voyelles, éléments formateurs de la signification lexicale et grammaticale, le dotent de sa substance-forme (*cf.* l'analyse en termes d'*exophrastie* et d'*endophrastie* dans Camoun 1992 : 179-180).

Ces précisions ne sont guère utiles pour le traducteur qui doit se livrer à un travail de description de la langue cible ; lequel travail serait fortement contraint du cadre conceptuel qu'institue la théorie.

Afin d'établir la cohérence recherchée entre l'appareillage conceptuel guillaumien et la langue décrite en ces termes, le traducteur linguiste se doit d'opérer à partir des configurations les plus générales de cette cohérence (fondées sur des ressemblances entre le français et l'arabe) à celles plus spécifiques qui viennent « perturber » la similitude illusoire entre les deux systèmes. En même temps, il se doit de situer ces spécificités dans le système général de la langue décrite, appréhendée à travers les concepts de la théorie.

Cette opération n'est pas sans risque d'autant que les différences entre les deux systèmes sont énormes. C'est ce que nous nous proposons de démontrer à partir des problèmes d'application à la langue arabe de quelques concepts clés de la théorie, à savoir ceux de l'incidence (appliqué à la distinction des parties de discours et plus particulièrement au substantif), du tenseur binaire radical (appliqué à un cas de la détermination) et de la chrongenèse (le verbe).

La notion d'incidence repose sur « la relation apport/support [qui] est un agent de la recatégorisation du mot. Un mot, en tout état de cause, est un apport de signification, et il se destine à un support. C'est en grande partie cette distinction et le mécanisme qu'elle s'attribue et se réserve qui font ce qu'on appelle la partie du discours – autrement dit, catégorisent le mot. » (Guillaume 1971 : 152) Mais, en tout cas, l'incidence n'est jamais achevée en langue. « La détermination de l'incidence formelle est reportée en discours. » C'est ce qui explique le lien établi par Guillaume entre la notion de prédicat (ce qu'on dit) et la notion d'incidence. « Avec la fonction de prédicat, nous sommes en présence d'un apport qui se recherche un support » (Guillaume 1973 : 69) et c'est sur cette base qu'il distingue les parties de discours prédictives « dont l'incidence interne [substantif], externe 1 [l'adjectif et le verbe], externe 2 [l'adverbe] ressortit à l'événement que la phrase relate » et les parties de discours non prédictives dont « l'incidence ressortit au mécanisme de l'événement qui est la phrase elle-même [la préposition] » (Guillaume 1982 : 130).

À un certain degré de généralité, c'est effectivement cette relation qui, en plus de l'incidence matérielle, donne lieu en arabe à une catégorisation tripartite du vocable arabe : le nom, le verbe et la particule.

En effet, du côté de la substance matière, les grammairiens arabes traditionnels avancent les définitions suivantes :

- « le nom est ce qui renvoie à une signification en lui-même, spécification abstraite au moment de l'intégration [dans le discours ; incidence interne] » (Azzamakhchari) ;
- « le verbe renvoie à la source et au temps du verbe » ;

- « la particule ne renvoie pas à une signification en elle-même ».

Du côté du discours, on retrouve ces définitions :

- « Le nom peut être support d'un prédicat et peut lui-même servir de prédicat » ;
- « le verbe est ce qui est prédicat et ne peut pas être support de prédicat » ;
- « la particule ne peut être ni support de prédicat, ni prédicat ».

Au-delà de cette catégorisation très générale, le traducteur se heurte à des sous-systèmes totalement étrangers à ceux analysés par Guillaume en français.

Si l'on prend le nom comme exemple, deux des problématiques que le traducteur-linguiste doit résoudre, sont l'intégration de l'espace qu'occupe l'adjectif en français dans la sphère du nom arabe et le fonctionnement de la détermination de l'indéfini à la défini [avec « Ø/al » ; ≈ « un »/« le »] en plus de ses affinités avec le verbe que nous traiterons dans le paragraphe portant sur la chronogenèse.

Tenant compte du traitement du nom dans la tradition grammaticale arabe, nous répartissons les douze types de nom (à peu près) allant du général au particulier et de la source au dérivé en trois classes définies selon la potentialité du nom à servir de support pour un autre nom :

a) Les noms ayant une incidence essentiellement interne :

- nom substantif, *ʔismu-lʔins* (primitif, commun, de genre) : d'une grande généralité ; *raʔulun* (homme) ;
- nom d'action, *ʔismu-lʔadaə* : source de la dérivation ; *naql* (transport) ;
- noms de lieu et de temps, *ʔisma: azzama:n wa-lmaka:n* : exprimant l'action au niveau de la langue et les circonstances de lieu et de temps dans le discours ; *manzil* (maison), *lajl* (nuit) ;
- nom d'instrument, *ʔismu-lʔa:la: maʔraqa* (marteau).

b) Les noms présentant un apport explicite et un support implicite susceptible d'être précisé dans le discours :

- nom-agent, *ʔismu-lfa:ʕil* : nom signifiant un qualificatif ; schème double signifiant l'action et son agent en même temps dans la même forme ; *na: qil* (transporteur) ;
- nom-patient, *ʔismu-lmafʕu:l* : qualificatif dérivé signifiant celui qui subit l'action et comporte la signification du qualifié et de du qualificatif dans la même structure ; *manqu:l* (transporté) ;
- qualité similaire, *ʕifa muʕabbaha* : forme dérivée d'un verbe intransitif ne signifiant pas l'action mais l'état, la continuité et l'abstraction ; *ʕari:h* (blessé) ;
- nom intensif, *ʔism muba:laya* : a la même signification que le nom-agent avec l'idée d'intensité ou de répétition ; *naqqa:l* (≈*très transporteur)

- l'élatif, *ʔismu-ttafdi:l* (nom de préférence): exprime le comparatif ou le superlatif et nécessite le plus souvent un support explicite; *anqil* (plus transporteur);
- nom relatif, *ʔism mansu:b*: genre de détermination (équivalent à N + N), exprime une qualité précise et un qualifié non totalement spécifié ou désigné; *tu:nisi*: (tunisien).

c) Le nom diminutif, *ʔism musayyir* qui tout en représentant le degré ultime de particularisation rejoint le nom substantif dans son incidence interne: apport spécifié et support spécifié; *bunajja* (fillette, petite fille).

Devant cette complexité, le traducteur doit:

- décrire, à travers des saisies différentes, le cinétisme permettant d'expliquer le passage de l'incidence interne (A), à l'incidence interne ou externe (B), à une nouvelle incidence interne (C) et situer ce cinétisme par rapport à l'opposition langue/discours;
- déterminer le processus d'effection du nom où se croisent trois niveaux:
- celui déterminé par le schème (*cf.* ci-dessus);
- celui de son fonctionnement dans le discours qui peut soit le basculer, d'une manière contingente, de l'incidence interne à l'incidence externe (A; *razulun* > *razulun ʔasad* = homme lion), soit fixer l'incidence dans l'un des deux versants (interne ou externe (B); *jubħaəu ʕani-lqa:til* = on cherche l'assassin; *wuʒida-rraʒulu-lqa:til* = *on a trouvé l'homme assassin);
- celui de l'extensivité déterminée par la présence ou l'absence de la détermination par *al* ($al \neq \emptyset$, $\approx le/un$).

Concernant ce dernier aspect relatif à la détermination, le traducteur ne peut se contenter des quelques similitudes entre le fonctionnement de l'article en français tel qu'il est décrit par Guillaume et l'opposition/« al » en arabe. Pour ce dernier, le cinétisme de l'article est régi par un psychomécanisme invariant de la pensée « se trouvant partout dans la structure des langues » qu'est le tenseur binaire radical. Celui-ci comprend deux tensions allant du large à l'étroit et de l'étroit au large. Dans le cas de l'article, le premier mouvement particularisant présente une orientation de l'universel au singulier (cinétisme de *un*); le second est un mouvement généralisant s'orientant du singulier à l'universel (cinétisme de *le*).

Dans la pensée grammaticale arabe, on retrouve quelques affinités avec cette manière de voir:

- « L'indéfini précède le défini. Initialement, le nom est indéfini (l'indéfini est rendu par *attanwi:n* à la fin du mot; *oun*, *an* ou *in* selon le cas) » (Ibn Siraj);
- « Le général précède le particulier »;
- « Certains indéfinis sont plus indéfinis que d'autres »

- La définitude par *al* peut référer soit à un singulier (surtout dans le cas d'une reprise du nom après son apparition sous sa forme indéfinie dans le discours), soit à toute la catégorie;
- etc.

Au-delà de ces affinités, le traducteur serait confronté à des difficultés lorsqu'il s'agit de définir, dans les deux mouvements descendant et ascendant, les saisies relatives à des cas comme :

- [1] *haða-lʔasadu muhi:ban* (littéralement : Ce le lion est imposant) où *alʔasad* (défini), précédé par le démonstratif *haða* réfère à toute la catégorie ;
- [2] *ka:na haða ʔams/alʔams* (litt. : Ceci était hier/*le hier) où le nom dans les deux cas (formellement indéfini et défini) est considéré comme défini.

Quant à la chronogenèse, telle qu'elle est définie ci-dessus, elle permet de se représenter la formation de l'image-temps linéairement (sur l'axe chronogénétique). Les saisies suspensives offrent en représentation :

- soit « une image temps à réaliser tout entière » (saisie initiale : mode quasi-nominal ; infinitif et participes en français) ;
- soit « une image temps partiellement réalisée » (saisie médiane : le subjonctif) ;
- soit « une image-temps complètement réalisée » (l'indicatif).

Si le système du temps demeure « égal à lui-même *in globo* » (Guillaume 1929 : 2), son architecture et sa construction diffèrent selon les oppositions propres à chaque langue.

En arabe, le système verbal s'organise autour de deux formes : une forme appelée *l'accompli* (F1) présentant une seule conjugaison à la troisième personne du singulier, une autre appelée *l'inaccompli* (F2) qui admet trois formes de conjugaison apparentes dans la voyelle suffixe :

- F2-u (*jaʕmal-u* ; souvent appelée par les arabisants *l'inaccompli indicatif*)
- F2-a (...*jaʕmal-a* ; appelée *subjonctif*)
- F2-° (...*jaʕmal* ; appelée *apocopé*).

L'emploi des qualificatifs *indicatif* et *subjonctif* laisse entendre une parenté avec le système verbal. Or, dans les faits, le linguiste ou le traducteur éprouvent des difficultés à ajuster les deux systèmes et surtout de rapprocher leurs descriptions dans les termes de la psychosystématique.

Avant même de soulever la différence fondamentale entre l'arabe et le français, on peut vérifier la non-pertinence de cette correspondance à partir des faits suivants :

- aux formes quasi nominales identifiées au niveau de la saisie initiale en français (infinitif et participes) correspond en arabe ou on cherche à faire correspondre soit la forme F1 à la troisième personne du singulier dont la valeur d'infinitif est pure fiction en arabe, soit les noms agents et patients (qu'on appelle souvent « participes »). Or, une équivalence a été souvent établie entre ces formes nominales et certains emplois des deux formes verbales F1 et F2. Ainsi, dans *ʔabi:ða:hibun ʔila:ssu:q* (litt. : *Mon père allant au souk*), *ða:hibun* équivaut à *jaðhabu* dans *jaðhabu ʔabi: ʔila:ssu:q* (litt. : *va mon père au souk*) et dans les deux cas, c'est le contexte qui se charge de l'expression du temps par l'adjonction par exemple du circonstant *aljawm* (« aujourd'hui »). De même, dans *haða-lʕamru mansijjun munðu sini:na* (litt. : *Cette affaire oubliée depuis des années*), *mansijjun* équivaut à la forme passive de l'accompli *nusijja* dans *nusijja-lʕamru munðu sini:na* ([on] *a oublié l'affaire depuis des années*). La valeur de l'accompli, appuyée par le circonstant temporel situe le procès dans le passé. En tout cas, on ne peut pas rattacher l'idée de « l'image-temps à réaliser en entier » aux formes nominales citées et en exclure les formes verbales;
- au niveau de la saisie médiane, le subjonctif en français émane en général d'une idée regardante virtualisante (souhait, volonté, etc.). Les deux formes en arabe sont susceptibles de rendre cette visée :

[3] Vive le roi : *ʕa:fa-lmaliku* (litt. : a vécu le roi ; F1)

li-jaʕifi-lmaliku (litt. : que vive le roi ; F2)

jaʕi:fa-lmaliku (litt. : vit le roi, F2).

- au niveau de la saisie finale, les deux formes sont susceptibles, dans des contextes précis, de situer le procès dans les trois époques :
 - l'inaccompli :

[4] *Lam jaðhab* *la: jaðhabu* *lan: jaðhaba*

N'est pas allé ne va pas n'ira pas

(remarquons l'intervention de la particule de négation dans le temps).

- l'accompli :

[5] *balaya maqsadahu bi-lʔamsi* (accompli dans le passé)

est arrivé à sa destination hier

[6] *balaya-lʔa:na maqsadahu* (accompli dans le présent)

est arrivé à sa destination en ce moment

[7] *saʔara:hu ʕindama: jaku:nu qad balaya maqsadahu* (accompli exprimant l'antériorité dans le futur)

je le verrai quand (il) est déjà est arrivé à sa destination.

En somme, on peut avancer que, contrairement au système verbal français où les marques temporelles se trouvent associées à la forme verbale, le verbe arabe ne porte pas de marque morphologique pour le temps. L'opposition entre les deux formes, en dehors de tout contexte, est de nature strictement aspectuelle et repose sur l'idée d'achèvement (F1) ou de non-achèvement (F2). Cependant, l'arabe est capable d'exprimer le temps à travers des particules et d'autres éléments contextuels. On peut bien penser à un système verbal où la particule détermine la forme verbale et où l'idée regardante qu'elle exprime détermine le mode. Tel n'est pas le cas. Dans les exemples suivants, la même particule est liée à des visées différentes ou à des formes différentes :

| | | |
|------------------------|----------------------------|-------------------------------|
| [8] <i>la: jaxruzu</i> | <i>la: jaxruʒ</i> | ... <i>la: jaxruʒa</i> |
| (il) ne sort pas | Qu'il ne sorte pas (ordre) | Qu'il ne sorte pas (souhait). |

De tout cela, nous ne voulons, en aucun cas, conclure que la langue arabe échappe à une description dans les termes de la théorie guillaumienne. Au contraire, nous avons tenté de montrer que le fonctionnement général de son système s'ajusterait aux principes généraux de la théorie. Mais, quand il s'agit de traduire cette théorie, on ne peut pas se passer de la description du système dans ses plus fins détails afin d'en dégager l'architecture qui lui est propre.

Conclusion

Traduire une théorie linguistique revient en fin de compte à transposer des termes et des concepts qui lui sont relatifs et à trouver des illustrations dans la langue cible pour ces mêmes concepts.

Au moment de la traduction, le traducteur se trouve face à deux types de difficultés. Au premier niveau, il doit tenir compte, d'une part, de la relation entre des traits conceptuels et un substrat lexical commun dont l'équivalence n'est pas toujours évidente lors du passage d'une langue à une autre, et d'autre part, du fait que ces traits s'intègrent dans des réseaux notionnels qui déterminent souvent leur sémantisme.

Au second niveau, trouver des illustrations dans la langue cible revient à une « réécriture » des fondements de la théorie à savoir la matière qui sert de support à la pensée conceptualisante. Autrement dit, il faudrait définir l'architecture de cette langue (avec ses structures et ses sous-systèmes) dans la perspective de cette théorie. Or, une telle tâche ne peut pas être concomitante au moment de la traduction : un travail théorique et descriptif préalable reste de toute évidence nécessaire.

Bibliographie

En arabe

- Achour (M.), 2004 : *Dhahiratu-ism fi-tafki: ri-nnahwi*, Publications de la Faculté des Lettres de la Manouba (Tunisie).
- Grira (T.), 2003 : *Almustalahu-nnahwi wa tafki: ri-nnuha: ti-lârab*, Publications de la Faculté des lettres de la Manouba (Tunisie).
- Ibn Alâkbari (B.), 1984 : *Charh-illamâ*, Le Koweit, V1 & 2.
- Ibn Jinni Abou-lfath (O.), 1952 : *Alkhsais*, Dar Alkitab Alâarabi, Beyrouth, 3 volumes.
- Ibn Siraj, 1985 : *Alousoul fi-nnahw*, Mouassasatu-rrisala, Syrie.
- Kchaou (S.), 1997 : *Madhahiri-ttaâri: f fi-lârabiyya*, Publications de la Faculté des lettres de Sfax (Tunisie).

En français

- Abi Aad (A.), 2001 : *Le système verbal de l'arabe comparé au français*, Paris, Maisonneuve & Larose.
- Blachère (R.), 2004 : *Grammaire de l'arabe classique*, Paris, Maisonneuve & Larose.
- Boone (A.), Joly (A.), 2004 : *Dictionnaire terminologique de la systématique du langage*, Paris, L'Harmattan.
- Ben Hamouda (R.), 2006 : « L'adjectif : classe de mots ou catégorie sémantique ? », dans *Actes des 6^{es} journées scientifiques tuniso-japonaises*, Sousse, novembre 2005, Mejri (S.), dir., Publication de la Faculté des lettres et des sciences humaines de Sousse, p. 15-22.
- Camoun (A.), 1992 : *Études de psycho-systématique française et arabe*, Publications de la Faculté des lettres de la Manouba (Tunisie).
- Chraïret (M.), 1996 : « Fonctionnement du système verbal en arabe et en français », dans *Linguistique contrastive et traduction*, numéro spécial, Paris, Ophrys.
- Guillaume (G.), 1929-1965 : *Temps et verbe*, Paris, Librairie ancienne Honoré Champion.
- Guillaume (G.), 1964 : *Langage et science du langage*, Librairie A.G., Paris, Nizet, Québec, Presses de l'Université Laval.
- Guillaume (G.), 1971a : « Structure sémiologique et structure psychique de la langue française 1 », dans *Leçons de linguistique 1948-1949*, série A, volume I, Paris, Klincksieck, Québec, Presses de l'Université Laval.
- Guillaume (G.), 1971b : « Psychosystématique du langage. Principes, méthodes et applications 1 », dans *Leçons de linguistique 1948-1949*, série B, volume II, Paris, Klincksieck, Québec, Presses de l'Université Laval.
- Guillaume (G.), 1973 : « Grammaire particulière du français et grammaire générale 4 », dans *Leçons de linguistique 1948-1949*, série C, volume III, Paris, Klincksieck, Québec, Presses de l'Université Laval.
- Guillaume (G.), 1974 : « Structure sémiologique et structure psychique de la langue française 2 », dans *Leçons de linguistique 1949-1950*, série A, volume IV, Paris, Klincksieck, Québec, Presses de l'Université Laval.
- Kouloughli (Dj.), 1994 : *Grammaire de l'arabe d'aujourd'hui*, Paris, Pocket.

- Mejri (S.), 1996 : « De la problématique de la construction de l'appareil terminologique linguistique arabe: spécimen méthodologique » (en arabe) [en collaboration avec (T.) Baccouche], dans *Études linguistiques*, volume I-1996, Publications de l'Association tunisienne de linguistique, p. 9-23.
- Mejri (S.), 1998 : « Le mot dans la tradition grammaticale arabe » [en collaboration avec (T.) Baccouche], dans *l'Information grammaticale*, numéro spécial *Tunisie*, mai 1998, Paris, p. 13-16.
- Mejri (S.), 2003 : « La traduction linguistique: problème terminologique ou construction conceptuelle? », dans *Traduire la langue, traduire la culture*, 4^e volume des *Actes du colloque Traduction humaine, traduction automatique, interprétation*, Maisonneuve et Larose, Paris, p. 177-191.
- Mejri (S.) [en collaboration avec Baccouche (T.)], 2006 : Traduction en arabe de *Pour une logique du sens* de Robert (M.), PUF, 2^e éd., 1992, Beyrouth, Publications de L'OAT.
- Mejri (S.) [en collaboration avec Baccouche (T.)], 2007 : « Norme grammaticale et description linguistique: le cas de l'arabe », dans Mejri (S.) et Neveu (F.), dir., *Langages* n° 167, numéro spécial sur la tradition grammaticale.
- Mejri (S.) [en collaboration avec Ouerhani (B.)], 2009 : Traduction en arabe de l'ouvrage de Gross (G.), *Les expressions figées en français*, Paris, Ophrys 1996, Publications du CERES.
- Mejri (S.), à paraître: Traduction en arabe du *Dictionnaire des sciences du langage* de (F.) Neveu, Armand Colin, 2004.
- Neveu (F.), 2004 : *Dictionnaire des sciences du langage*, Paris, Armand Colin.
- Ouerhani (B.), 2006 : « La traduction de la métalangue: la problématique *terme/mot* en contexte », dans *Actes des 7^{es} Journées scientifiques du réseau LTT*, Bruxelles 8-9-10 septembre 2005, *Actualités scientifiques*.
- Saïd (M.), 2006 : « Traduction et métalangue: de la typologie de l'adverbe en français aux typologies des équivalents en arabe », dans *Actes des journées scientifiques La terminologie linguistique: problèmes traductionnels, conceptuels, épistémologiques*; 12/13 mai 2005, Université de Caen. *Syntaxe et Sémantique*7, p. 153-168.
- Saïd (M.), 2007 : « La transposition de la terminologie grammaticale: les procédés et leurs impacts », dans *À la croisée des mots, hommage à T. Baccouche*, Publications de la Faculté des lettres de Sousse.

Lucie Raharinirina Rabaovololona,
DIFP-Université d'Antananarivo,
Antananarivo,
Madagascar,
rabaovololona@moov.mg.

Traduction des expressions à emprunt vers leur langue source, un rite de passage de la diversité linguistique et culturelle

Résumé

Le malgache emprunte aux langues véhiculaires, notamment le français. Intégrés totalement ou partiellement dans une autre communauté, ces emprunts ont un nouveau statut dans le système intégrateur. En témoigne leur traduction vers la langue « génitrice » (sous sa forme standard et/ou régionale). Des expressions à emprunt de l'art culinaire, extraits de textes bilingues écrits (menu, carte, annonce publicitaire en tourisme) et oraux (rituel de commande au restaurant) illustrent autant leur changement dans le contexte lexico-sémantique malgache que les impacts dans l'espace commun qu'est la francophonie. L'instrumentalisation de leur traduction par des outils interactifs aidera à ce passage.

Mots-clés: expressions à emprunt; traduction; diversité linguistique; diversité culturelle.

Introduction

Comme toutes les langues locales des îles indianocéaniques, le malgache a vu l'intégration dans son lexique d'emprunts aux langues de grande expansion, notamment le français. Le contexte actuel marqué par la globalisation synonyme d'échanges, mais plutôt orientés du centre vers la périphérie et avec sa prolifération de termes techniques, ne fait qu'accentuer ce statut satellitaire d'une langue non véhiculaire. Perçus comme des éléments extérieurs au système, et de compréhension inhérente à leur langue d'origine plutôt qu'à celle emprunteuse, les emprunts ont souvent été exclus du pré carré à préserver de l'espace linguistique local des pays anciennement colonisés comme Madagascar. Dans la littérature sur les enjeux linguistiques, on les catégorise comme des points forts d'évaluation d'une concurrence entre langue dominante et langue dominée.

Un regard manichéen, qu'il y ait recours ou non aux emprunts, ne saurait cependant pas permettre un juste retour des choses qui reconsidérerait ces éléments de recours dans leurs fonctions primitives: le développement quantitatif de la langue emprunteuse et le dialogue des cultures à travers ces éléments « d'ailleurs ». Aussi,

la question que nous aborderons est plutôt celle de leur nouveau statut dans le système intégrateur, et ce, à travers un phénomène le moins attendu : leur traduction vers la langue « génitrice ». La finalité de l'étude est une perspective d'exploitation en lexicographie d'apprentissage (Conceicao 2004) avec des liens entre des textes bilingues ouverts appropriés et une bibliothèque virtuelle constituée d'outils interactifs (liens hypertextes et formes évolutives).

Mais avant l'exposé proprement dit, un bref rappel sur l'emprunt s'impose. L'emprunt fait partie des moyens dont disposent les locuteurs pour accroître leur lexique. Le processus consiste, pour une langue, à introduire dans ce lexique un terme venu d'une autre langue. Les langues empruntent surtout des mots appartenant aux classes lexicales « ouvertes », c'est-à-dire justement celles qui contiennent un stock variable de lemmes comme les noms, les verbes et les adjectifs. À l'arrivée, il n'est pas rare que la langue emprunteuse change ces catégories initiales, par exemple, un nom devient adjectif, un syntagme est pris comme un nom (*la bière* transformée en un seul lemme *labiera*). Cette assimilation peut dépasser l'adaptation phonique qui marque un transfert notionnel comme dans *karaoty* « carotte » et être doublée d'autre(s) changement(s) comme celui morphologique de *ladobo* venant du groupe nominal *la daube*, mais transformé en un seul élément grammatical non analysable morphologiquement : un nom pouvant être déterminé (*ny + ilay*) *ladobo*, « (la + cette) daube ». Ces emprunts servent de repères aux non-connaisseurs de la langue pour la compréhension sommaire d'une conversation. Lors d'une commande au restaurant, un francophone passant derrière un malgachophone saisirait au moins le contenu de sa commande avec de tels emprunts :

[1] *Anaovy salady karaoty, azafady ô!* « Une salade de carotte s'il vous plaît ! »

Cependant, l'examen des faits nous a guidée vers des phénomènes plus complexes. En effet, en usage dans des communautés différentes de leur communauté d'origine, ces lexèmes font partie du processus de développement des deux langues en contact : les mots empruntés évoluent au contact de la langue d'arrivée si bien qu'ils s'écartent du sens originel et mènent vers la diversité linguistique tout en constituant en eux-mêmes un médium dans l'espace culturel francophone pluriel. C'est en ce sens que nous associons leur traduction à un rite de passage qui marque le changement de statut social d'individus, d'un groupe à un autre, passage que (van Genep 1981) définit comme « l'interface et le paradoxe de la superposition partielle dans l'intersection de deux univers habituellement disjoints de l'agrégation et de la séparation à la fois ». Ces univers sont ici d'ordre linguistique : d'une part, le malgache, langue nationale non véhiculaire emprunteuse au français, et, d'autre part, l'espace francophone qui utilise à la fois la forme standard et la variante régionale du français.

Pour mener l'étude, nous avons rassemblé un corpus constitué de textes bilingues écrits (menu du jour, carte de restaurant, annonce publicitaire) et/ou oraux (rituel de commande au restaurant...); on a aussi intégré des extraits de copies d'étudiants en CLAT (Communication et Langues appliquées au Tourisme). Pour interpréter

les faits, notre expérience de linguiste et de lexicographe a été conjuguée avec un regard transversal de traduction interculturelle, d'ailleurs, nous nous inscrivons dans une perspective plus culturelle que linguistique. Les outils méthodologiques fournis par les classes d'objets de Gaston Gross et son équipe permettent en effet de faire une étude lexico-sémantique qui valorise les diverses combinaisons d'un lexème de base, sans privilégier le statut du locuteur. C'est le regard contrastif ou confrontation des formes francophones et malgaches qui nous a menée vers les contraintes lexico-syntaxiques de leur nouveau système.

Un début d'analyse des résultats va être présenté en trois temps : les apports des expressions à emprunt du français dans le contexte lexico-sémantique malgache, leurs impacts dans l'espace commun qu'est la francophonie, et l'instrumentalisation par le projet de cours de traductologie en ligne.

1 Les expressions à emprunt du français en contexte lexico-sémantique malgache

Plutôt introduits pour pallier des carences d'ordre conceptuel, les emprunts sont considérés comme relevant de leur langue d'origine et les soucis des utilisateurs et des grammairiens dont (Raharinirina Rabaovololona, Ralalaoherivon, Ranaivoson 2001) reposent souvent sur leur intégration orthographique. Cependant, au cours de nos travaux de traduction spécialisée, notamment en tourisme, il est apparu que certains emprunts ont subi une adaptation sémantique. Cela a pour conséquence en traduction vers la langue source des contresens et même des non-sens. Le mot emprunté est, en effet, souvent arrivé vierge de ses connotations, celles-ci étant des représentations et des symboles du culturel de chaque communauté linguistique. En témoigne leur re-passage vers le français qui montre de nouvelles propriétés, donc nouveaux sens, par rapport aux sources françaises. L'exemple de *hôtel* devenu *hotely* en donne l'illustration. Si Le *Petit Larousse* (2008) le définit comme « un établissement commercial qui loue des chambres ou des appartements meublés pour un prix journalier », l'emprunt présente trois sens (dont deux innovations locales) justifiés par des combinaisons lexico-syntaxiques différentes. Si on a la construction de même sens dans :

[2] *Matory amin'ny hotely*, « dormir à l'hôtel » (V Prép Dét déf N),

on a un autre emploi lié au sens vernaculaire « auberge », mais nécessitant un modificateur adjectival qui en fait une locution opposée à *restaurant* « établissement » dans :

[3] *Mihinana amina hotely gasy* « manger au « coin », manger local ».

L'emprunt donne aussi une expression figée spécifique qui demande des compétences multiculturelles au traducteur :

[4] *Manao hotely ny tranon'olona*, « entrer dans la maison de quelqu'un comme dans un moulin ».

Cependant, le choix sémantique utilisé dans cette métaphore relève quand même du même concept de *libre accès*. À noter aussi que le malgache possède une variante locale pour certaines de ces formes, mais l'intégration de l'emprunt a développé quantitativement le choix conceptuel et/ou lexématique des locuteurs, ce qui est important pour les enjeux du multilinguisme et l'interculturalité. Outre ces dégroupements de sens, on notera trois types d'innovation locale :

1.1 L'emprunt conserve son sens premier sens mais s'utilise dans des collocations et expressions locales

Si le mot simple *labiera* est le correspondant du français *bière*, les collocations obtenues demandent une recherche d'équivalence qui dépasse le concept aristotélien de la traduction : *labiera mangatsiaka* littéralement « bière froide » désigne *bière glacée* et *labiera mafana*, « bière chaude » se traduit par *bière non glacée*. Le malgache comporte encore des néologismes par dérivation nominale agentive comme dans l'expression *mpandabiera gaigilahy*, « grand *biéreur ». Il est clair que le correspondant n'existe pas dans la langue source et que le traducteur se doit de chercher dans les expressions et collocations : *gros buveur de bière*, *amateur de bière*. Ainsi, on peut dire que ces formes à innovation sémantique confirment l'appropriation de l'emprunt par la langue locale.

1.2 L'emprunt ne conserve que des liens métonymiques avec des formes opaques

La création terminologique par métonymie est un procédé assez courant (Lino 2001) ; le sens original passe souvent du particulier au général comme dans l'emprunt *dite* (à lier à *thé*) qui désigne *toute boisson à base d'eau bouillie à vocation désaltérante et/ou médicamenteuse* : *tisane*, *infusion*, *décoction*... Les expressions qu'on obtient justifient une différenciation culturelle entre les deux espaces : *dite vazaha*, littéralement *thé étranger* désigne le terme source *thé*. Certaines réalités dénotées avec le lexème peuvent même créer des ambiguïtés au sein du français, langue source : *dite veromanitra*, « thé citronnelle » désigne et l'infusion et le thé à la citronnelle (mélange exotique).

1.3 L'emprunt développe des spécificités distributionnelles

L'enrichissement de la langue locale par intermédiaire de l'emprunt peut être d'ordre syntaxique ; des structures collocationnelles d'intensif comme *mihinan-kafe*, « manger du café », « être friand de café », ou *mihinan-divay*, « manger du vin », « aimer le vin », deviennent courantes. D'autres combinaisons ajustées par le système naturel se révèlent dans la langue emprunteuse, si bien qu'il devient impossible d'imiter la structure autre que phonique de l'original. Comme explication de ces changements morpho-syntaxiques, le changement d'emploi qu'en fait le malgache, or « un emploi est d'abord défini par le schéma ou domaine d'arguments du lexème » (Gross 1994).

Exemples :

- [5] *Sotron-dite*, « cuillère à thé », pour cuillère à café ;
 [6] *Kafe katsaka*, « café maïs », « café allongé », « café doux » (trop dilué, au point que la boisson n'a de café que l'arôme et la couleur ; on assimile cette boisson au jus du maïs nature cuit à l'eau) par opposition à *kafe mahery*, « café fort ».

Dans ce dernier exemple, la mise en relation des deux concepts (*café* et *maïs*), quoique tous deux végétaux, ne s'explique que par la différence de l'environnement culturel et des pratiques liées à chaque langue. La métaphore de *l'eau de maïs* ne peut être associée à *café* du français et n'est pas traduisible en soi non plus car la forme constitue une spécificité de l'emprunt *kafe* en contexte malgache. Le passage entre les deux contextes (emprunteur/prêteur) est primordial pour l'interculturalité et le multilinguisme. D'ailleurs, l'Unesco, dans sa recommandation de 1996, a déjà insisté sur le rôle plus qu'important de la maîtrise de la langue nationale pour aborder une seconde langue et la pratiquer de façon systématique comme dans les pratiques traductionnelles. Si tels sont les faits afférents au phénomène d'emprunt en malgache, en français on constate aussi un retour non négligeable de ce nouvel usage.

2 La traduction des emprunts vers l'espace commun qu'est la francophonie

À partir d'analyse des propriétés de l'emprunt, notamment les traits sémantiques et les spécificités distributionnelles, nous verrons apparaître quelques aspects divergents entre le lexème français et celui emprunté par le malgache ; la connaissance de ces propriétés conditionne leur traduction en français.

Le français local voit ainsi émerger des néologismes, issus de pratiques culturelles spécifiques du contexte malgache. Comme exemple le mot valise *thécao*, « boisson chaude au cacao », qui s'utilise à la place de *chocolat chaud*. Ces innovations ont quelquefois un impact ontologique ; en effet, le *cresson* de l'expression *bœuf-brède cresson*, « zébu mélangé avec des feuilles de cresson », n'est plus classé dans les légumes mais dans les *brèdes*, classe de végétal comestible mais non répertorié dans la classe des *légumes* dans le domaine de la gastronomie malgache. Or, un emploi est déterminé, au moins en partie, par le domaine dont il relève, donc le phénomène mérite une investigation si on veut arriver à une opérationnalité des outils translangagiers du tourisme. Ce fait est d'ailleurs systématiquement utilisé dans les différents systèmes de traduction automatique. Ces derniers doivent en effet noter le domaine de spécialité utilisé dans la ou les langues à traduire car cela contribue à déterminer le choix de telle ou telle traduction parmi les différentes possibilités.

Ce transfert de classe peut être doublé d'un élargissement de sens, ou même d'une rupture avec le contexte sémantique de départ ; cela est concrétisé par une autre combinatoire, comme le verbe approprié. Par exemple, l'expression malgache *manao*

laoka lasopy, « faire un mets soupe », « prendre un bœuf aux légumes » comporte l'emprunt *lasopy*, qui vient du français *soupe* défini dans le *Larousse* (2008) comme *potage* ou *bouillon*, et désignant soit un plat unique de la classe des entrées soit le nom générique *nourriture*. En malgache, l'emprunt *lasopy* signifie par contre un mets de la classe des plats de résistance accompagnant le riz, une traduction culturelle donnerait plutôt *bœuf aux légumes* sans qu'on soit satisfait du résultat puisque seuls les ingrédients sont rapportés dans le terme, la modalité reste à expliciter : *cuisson avec jus, conçu pour accompagner le riz sec*.

Pour les phraséologues, souvent l'accès à l'information collocationnelle se fait à partir de la base; si la traduction de l'emprunt dans la langue source ne ramène plus nécessairement au même lexème comme ci-après, la traduction demande une investigation culturelle à faciliter par les outils TIC. Pour concrétiser, nous allons récapituler dans un tableau les informations sémantiques sur le même emprunt *lasopy* et ses liens avec le lexème *soupe*. Le contexte d'usage est précisé avec la traduction pour bien montrer la diversité constatée. Nous symboliserons les sens français par *Sf* et ceux du malgache par *Sm*.

| SOUPE <i>Sf</i> | LASOPY <i>Sm1 = Sf</i> | TRADUCTION DE L'EMPRUNT |
|--|--|--|
| Bouillon ou potage épaissi par des tranches de pain, des aliments solides (généralement non passés comme dans la soupe au pain, la soupe aux choux, aux pommes de terre. | <i>Karazan-tsakafo tsy misy ventiny firy fa ny ro no tena ilana azy.</i> Ex. : <i>Lasopy ihany izao no mbola misy ato fa tsy masaka ny vary.</i> « On n'a que de la soupe pour le moment; le riz n'est pas prêt ». | « Soupe » (un menu ou une partie du menu). |
| | <i>Sm₂ =</i> <i>Henomby sy legioma isan-karazany be rony fanao laoka</i> Ex. : <i>Anana foana ve no fangaron-daoka fa mba manaova laoka lasopy indray e</i> « Encore des brèdes pour accompagner la viande ? Faut faire du bœuf-légumes ! » | « Bœuf-légumes » (un accompagnement du plat de résistance, généralement constitué par du riz sec). |
| | <i>Sm₃ =</i> <i>Sakafo maivana azo hanina alohan'ny tena sakafo hariva</i> Ex. : <i>Avy eo mihaona de mba miara-mihinandasy any e!</i> « À + et on en profitera pour prendre un en-cas ensemble ! » | « Encas », goûter pour grandes personnes et qu'on peut retrouver dans l'expression <i>prendre un petit quelque chose</i> . |

3 L'instrumentalisation par le projet de cours de traductologie en ligne

Les phénomènes sus exposés montrent les difficultés de traduction entre le français et le malgache, car il faut réinterpréter les emprunts, non pas par rapport à leur source mais en termes usités dans leur nouveau contexte, *au risque d'en déformer le sens* (Rajaspera 2001). La déformation est un indicateur de multiculturalité que toute réflexion didactique pour un bilinguisme additif doit prendre en compte. On ne peut traiter d'erreur sémantique un néologisme de fait provoqué par l'intégration de l'emprunt. Dorénavant, « les réponses structurelles interpellent prioritairement l'engagement plus créatif des chercheurs pour associer les praticiens de l'éducation vers une convergence des recherches... L'ouverture aux langues, aujourd'hui, exige de nouveaux paradigmes pour un partenariat des langues, une didactique qui doit se départir d'un conservatisme inopérant » (Rakotoanosy 2007). Dans la lignée de cette réflexion, nous avons la perspective de valoriser la recherche sur la traduction des emprunts par une exploitation d'outils interactifs, capitalisation de nos expériences avec les équipes de recherche partenaires, notamment le réseau des lexico-grammairiens et l'équipe de Paule Fioux de la Maison des langues de l'Université de la Réunion, Campus Sud. En plus, de la mise à disposition d'outils plus pratiques, notre projet est conforme aux questionnements de (Raharirinirina Rabaovololona et Ralalaoherivony 2007) sur comment « résorber une situation pour le moins complexe de didactique divergente entre des langues qui cohabitent » dans l'espace francophone.

3.1 Actualisation et mise en version numérique du LSM avec des liens aux dictionnaires monolingues français de référence (*Robert, Larousse-Lexis*)

Le LSM est un lexique terminologique du tourisme local en cinq langues, réalisé par le groupe LATA (Langues appliquées au tourisme et aux affaires) créé en 1999 au sein du DIFP de la FLSH de l'Université d'Antananarivo. Dans une perspective didactique de lexicographie d'apprentissage, il est essentiel que le dictionnaire fournisse une information linguistique détaillée sur les collocations, si on veut que les étudiants les emploient à bon escient. Ainsi, l'information syntaxique doit préciser le type de constructions accompagnant le mot ainsi que les aspects distributionnels comportant les mots appropriés au sens décrit. Si dans la version actuelle, on a présenté dans la rubrique des aliments *kompoze* qui est un emprunt du mot *composé* (nom générique désignant un ensemble formé par plusieurs éléments ou parties) mais se traduisant par « assiette de crudités, de macédoines et/ou de nouilles en salade », la forme actualisée mettra en exergue la néologie syntaxique *mihinana kompoze* qui n'a pas de correspondant *manger un composé*. Même cas pour *manao laoka karatoa* comportant l'emprunt de *grattons*, « résidus de la fonte de graisse de volaille qu'on prend froid », mais dont le contexte n'est plus similaire puisque *karatoa* est cette fois un résidu de viande de porc pris (chaud ou froid) avec le plat de riz sec. Ce qui nous obligerait à mener à une actualisation qui nécessiterait une approche théorique plus ajustée : les classes d'objets et leurs exploitations en lexicographie multilingue (Gross 1992).

3.2 Mise en contexte des emprunts par des textes multilingues

Le problème (ou la problématique même) de la compétence linguistique des jeunes universitaires, aussi bien en malgache qu'en français, a été noté par plusieurs chercheurs en langues. Notre expérience dans le projet régional SDL-AUF *Pratiques et perceptions du français par les étudiants : étude des cas des promotions entrantes dans la perspective d'une francophonie universitaire s'articulant avec les langues nationales* nous a permis d'avoir une vision sur les possibilités offertes par les outils TIC dans la recherche d'une solution universitaire au problème évoqué. Le choix de Moodle, qui est une plate-forme d'apprentissage en ligne sous licence *open source*, promeut la création de communautés d'apprenants autour de contenus et d'activités pédagogiques et peut aller dans le sens d'une future mise en commun entre les partenaires de l'océan Indien (à noter la mise en place actuelle d'une plateforme de chercheurs sur le plurilinguisme des universités malgaches dénommée SLANDI (société, langues et diversité)). En effet, ces entités ont pour perspectives de constituer une base de données de connaissances sur la question des langues et aussi de mettre en commun des supports de formation en matière de didactique, dont la terminodidactique dont il est question ici. Le profil technique de Moodle – qui, hormis le système de gestion de contenu (SGC), ajoute des fonctions pédagogiques ou communicatives pour créer un environnement d'apprentissage en ligne – nous semble favoriser des interactions entre enseignants et apprenants en traduction d'une part et les ressources pédagogiques d'autre part, notamment par le système de bibliothèque virtuelle. Une partie de celle-ci sera alimentée par des textes parallèles comportant les expressions à emprunts pour pouvoir accéder aux exemples et expressions par la recherche en texte intégral qui balaie tous les articles. Le cours de traductologie exploitera ces éléments par des granularités de grammaires locales, traitées par le lexique-grammaire et les classes d'objets.

3.3 Veille terminologique

Deux dimensions sont à promouvoir dans cette activité de veille : le national et le régional.

Au plan national, à Madagascar, les écrits font défaut du fait que jusqu'ici les traditions orales priment encore. Cependant, nous aurons les possibilités de mettre cette plateforme en lien avec d'autres sites issus de projets tutorés de nos étudiants en communication, comme celui sur le tourisme national ou les cultures malgaches. Ce site interdisciplinaire sera particulièrement enrichi avec les mémoires universitaires, les rapports de projets tutorés et stages ou des extraits de textes médiatiques. Le parachèvement de l'outil en banque de données terminologiques avec rubrique sur les emprunts, ne sera assuré que s'il y a une intégration des usages en langue orale, malgré la difficulté de leur validation. Par *texte*, nous désignons en effet les formes normalisées et celles non structurées qui sont quand même témoins de l'usage et des pratiques de l'emprunt. Cette initiative peut aussi être associée à la pratique de traduction spécialisée malgache-français où on a des difficultés à trouver des textes assez représentatifs du culturel lié au tourisme.

Considérons à présent le plan régional. Une recherche terminologique n'a de valeur que dans la mesure où ses résultats sont diffusés en vue d'une éventuelle validation. Le phénomène d'évolution sémantique des emprunts n'est pas spécifique au malgache ; cela peut se retrouver aussi entre deux langues véhiculaires comme le français et l'anglais où l'on ne partage pas les coutumes ni le mode de vie (Soubrier 1999). Aussi, vu le contexte indianocéanique où l'on retrouve des usages convergents d'emprunts, nous estimons qu'une mise en réseau sera bénéfique à l'espace francophone. Comme illustration, le lexème *soupe* sus-étudié est emprunté par le créole mauricien sous la forme *la-soupe* attestée dans *li pas content la-soupe légitime*, « elle n'aime pas la soupe aux légumes » ou encore *bôl la-soupe* désignant la « soupière ». Le comorien ou *shi komori* réalise ce mot en *sopo* et a des usages distributionnels innovants. Le projet devra donc connaître une perspective régionale autant pour le Réseau LTT que pour les enseignants de traduction et de langues.

Conclusion

La compréhension cognitive nécessaire à la traduction des emprunts vers leur langue source demande la conjugaison de *connaissances conceptuelles* en langue emprunteuse avec les *connaissances notionnelles* d'origine. Le contexte verbal bilingue devrait pouvoir faciliter cette mise en lien, si on fait une mise en lien du discours avec la langue. Utiliser des moyens interactifs serait alors un rite de passage vers l'interculturalité, forme optimale de l'intercompréhension entre diversités linguistique et culturelle.

Pour terminer, nos remerciements à :

- Réseau LTT de L'AUF, pour les 8^{es} Journées scientifiques ;
- BOI-AUF, pour la mise en réseau des chercheurs en langues de l'océan Indien qui garantit la suite de ce projet ;
- Université d'Antananarivo, avec l'appui institutionnel à la recherche en langues ;
- FLSH et DIFP qui hébergent le CERCOM (Centre de recherche en communication), adossé à l'UFR CODYL options Communication et Langues Spécialisées (CLAS).

Bibliographie

- Andrianierenana (C.) *et al.* en collaboration avec Le Pesant (D.) et Vives (R.), 2000 : « Problèmes de traduction des expressions figées : un exemple bilingue français-malgache, malgache-français », dans *Actes du Colloque International Traduction humaine, traduction automatique, interprétation* tenu à Tunis du 28 au 30 septembre 2000.
- Andrianierenana (C.) *et al.*, 2001 : *Lexique Spécifique Malgache du Tourisme*, Antananarivo, Ev Edition.
- Andrianierenana (C.) *et al.*, 2002 : *Lexique International de Base du Tourisme*, Ev Edition Antananarivo.
- Conceição (C.M.), 2004 : « Concepts et dénominations : Reformulations et description lexicographique d'apprentissage », dans *Études de linguistique appliquée*, n° 135, 2004/3.

- Gross (G.), 1992 : « Formes d'un dictionnaire électronique », dans Clas (A.) et Safar (H.) *L'environnement traductionnel*, Presses de l'Université de Québec, Sillery Canada, p. 255-271.
- Gross (G.), 1994 : « Classes d'objets et description des verbes », dans *Langages*, n° 115, Larousse, Paris.
- Gross (M.), 1981 : « Les bases empiriques de la notion de prédicat sémantique », dans *Langages*, n° 63, Larousse, Paris.
- Lino (M.T.), 2001 : « De la néologie à la lexicographie spécialisée d'apprentissage », dans *Cahiers de Lexicologie*, n° 78, Paris, Honoré Champion.
- Raharinirina Rabaovololona (L.) et Ralalaoherivony (B. S.), 2003 : « Traduction français-malgache et réduction des distances », dans *Actes du colloque Traduction en francophonie*, Rennes.
- Raharinirina Rabaovololona (L.), 2006 : « Traduction et médiation, les incontournables de l'intercommunautaire : cas du lexique foncier franco-malgache dans la sécurisation foncière à Madagascar », Communication au colloque, *Traduction et communautés*, Université de Bretagne Sud, Lorient, France.
- Raharinirina Rabaovololona (L.), 2008 : « De l'insécurité linguistique à l'insécurité de l'espace public ou les enjeux d'une déconstruction linguistique pour le développement de Madagascar », dans *Actes du colloque Les mots du développement : genèse, usages et trajectoires*, Paris.
- Raharinirina Rabaovololona (L.), Ralalaoherivony (B.S.); Ranaivoson (J.F.), 2001 : « Fitsipi-panorantana », Premier numéro de la revue interne *Teny malagasy* du CIRAM – DLLM, FLSH Antananarivo : Imprimerie de la RN7, 40 pages.
- Rajaspera (R.), 2001 : « L'équivalence dans la différence : l'énonciation juridique d'une langue à l'autre », dans *L'éloge de la différence : la voix de l'autre. VI^e Journées scientifiques du Réseau thématique Lexicologie, Terminologie, Traduction*, Beyrouth, Publications des Universités Francophones AUPELF-UREF, Paris.
- Rakotoanosy (M.), 2007 : « Appropriation du français : problématique et perspectives à Madagascar », dans *Actualités scientifiques, Rencontres régionales ODFLN-SDL de l'île Maurice*.
- Ralalaoherivony (B. S.) et Raharinirina Rabaovololona (L.), 2007 : « Plurilinguisme à Madagascar Quel statut pour le malgache ? », dans *Actes du colloque Plurilinguisme à Madagascar, dans l'océan Indien et au-delà. Quand les perspectives se rencontrent. Situation plurielle. Description et analyse de la situation*, ENS Université d'Antananarivo, Madagascar.
- Ralalaoherivony (B.S.), 2004 : *Création d'outils méthodologiques en vue de la modernisation du malgache*, synthèse des travaux présentés pour la HDR, Université d'Antsirananana.
- Soubrier (J.), 1999 : « Néologismes et termes d'emprunt dans le français des affaires », dans *Actes du colloque La mémoire des mots, V^{es} journées scientifiques du réseau thématique Lexicologie, terminologie, traduction, Université de Tunis, 25-27 septembre 1997*, Tunis, AUPELF-UREF, Editions AUPELF-UREF, Collection Actualité Scientifique, p. 403-418.
- van Gennep (A.), 1981 : *Les rites de passage. Étude systématique des rites*, A. J. Picard, Paris.

André Valli,
Département de linguistique française,
Université de Provence (Aix-Marseille 1),
Aix-en-Provence,
France,
André.Valli@univ-provence.fr.

Alberto Bramati,
Dipartimento di Lingue e Culture contemporanee,
Università degli Studi di Milano,
Milano,
Italie,
alberto.bramati@unimi.it.

Traduire les métaphores ontologiques dans l'espace roman. Le cas du français vers l'italien

Résumé

Nous nous intéressons aux constructions verbales figées en français et en italien contenant un nom « partie du corps », comme *avoir du nez*, qui s'organisent autour d'un réseau de métaphores ontologiques (Lakoff et Johnson 1980).

Dans une première partie, nous abordons le problème posé par ces expressions aux théories linguistiques : déviance de la locution/systématicité de la langue (non-substituabilité, non-permutabilité, double lecture, non-prédictibilité).

Dans la deuxième partie, d'abord nous présentons un corpus français de constructions verbales figées contenant *œil* et *nez* ; ensuite, nous analysons les relations existant entre ces constructions et leurs expressions correspondantes en italien ; enfin, nous montrons dans quels cas la traduction en italien paraît prévisible.

Mots-clés : métaphores ontologiques ; constructions verbales figées ; noms « partie du corps » ; français ; italien.

Introduction

On connaît l'intérêt suscité, en particulier dans les rencontres organisées par l'AUF, par les travaux qui s'intéressent aux situations de plurilinguisme qui peuvent se créer entre langues proches, génétiquement apparentées et culturellement voisines. Une expérience d'apprentissage de la lecture en langues romanes étrangères comme Eurom4 (Blanche-Benveniste et Valli 1997), si elle ne débouchait pas sur

un programme d'apprentissage fédéré de ces langues, attirait fortement l'attention sur les processus d'intercompréhension dans un contexte plurilingue. Cette expérience a montré, prenant le contre-pied des approches traditionnelles tablant sur la singularité de chaque langue, qu'il existe entre les langues romanes infiniment plus d'homologies de structure que de différences. Finalement, on s'aperçoit au cours de ce type d'apprentissage, que seul un minimum de connaissances est nécessaire pour permettre une intercompréhension de qualité acceptable dans le processus de lecture en langue romane étrangère. On en perçoit tout l'intérêt : cette capacité acquise à lire en langue romane étrangère rendrait superflu le recours à une traduction en anglais de textes rédigés initialement dans chacune de ces langues.

Dans cette communication, nous nous intéressons aux constructions verbales figées en français et en italien contenant un nom « partie du corps », comme *Mener qqn par le bout du nez*, en nous situant toujours dans la perspective de la recherche des conditions d'une intercompréhension romane. Les travaux de Lakoff et Johnson (1980) et plus récemment de Lakoff (1987) ont attiré l'attention des linguistes sur le rôle primordial de la métaphore dans la mise en place de nos conceptualisations. Ces travaux sont ainsi à l'origine d'un regain d'intérêt pour l'étude des locutions figées, en particulier les expressions contenant un nom « partie du corps », qui déclenchent, selon F. Bercker (1992 : 126), « des égalités qui permettent de structurer notre expérience d'une chose en termes d'autre chose, d'employer des concepts moins complexes relevant de notre perception/interaction avec *moi + les autres + le monde*, pour exprimer des concepts plus complexes ou abstraits ». Bercker fait en outre remarquer que l'ensemble des constructions figées concernant les parties du corps « semble s'organiser autour d'un réseau de métaphores qui dépendent toutes à divers degrés, de ce que Lakoff et Johnson appellent la *métaphore ontologique* qui consiste à concevoir une notion concrète ou abstraite comme une entité, objet ou substance » (1992 : 126).

Le problème crucial que posent ces expressions est que, par définition, elles sont figées – leur construction est fixe – et d'autre part, leur sens n'est pas compositionnel, ce qui devrait soulever, *a priori*, un problème spécifique d'intercompréhension et donc une difficulté de traduction, SAUF à poser que les Européens de langue romane auraient tendance à développer des processus de conceptualisation de leur expérience du monde sensible relativement proches. Nous avons adopté cette hypothèse dans une précédente étude (Valli et Villagenes-Serra 1998) portant sur les métaphores ontologiques de l'espagnol et du français qui avait abouti à une confrontation intéressante entre les deux listes de constructions verbales figées dégageant des cas fréquents d'homologies totales ou partielles. Ainsi, les structures grammaticales sont identiques modulo une différence de déterminant, de modificateur, de préposition ou de verbe. Nous la poursuivons ici en nous proposant de confronter deux listes de constructions verbales figées contenant un nom « partie du corps » en français et en italien.

1 Description linguistique des métaphores ontologiques françaises mettant en jeu un nom « partie du corps »

Les travaux sur la locution qu'a réunis Martins-Baltar (1977) se sont intéressés aux problèmes théoriques auxquels sont confrontés les linguistes décrivant les usages observés sur corpus du fait de cette extension récente du domaine de la locution avant tout dans sa dimension intralinguistique (l'expression des lieux communs, des idées reçues, les métaphores de la vie quotidienne). Il s'agit essentiellement du problème posé par la locution aux théories linguistiques : déviance de la locution/systématicité de la langue (non-substituabilité, non-permutabilité, double lecture, non-prédictibilité). C'est la raison pour laquelle le cadre théorique de cette étude est emprunté principalement aux travaux de Maurice Gross et Laurence Danlos.

1.1 Un point sur la définition et la description des constructions figées en français

Danlos (1988) définissait le figement par l'opacité du sens et la fixité des éléments, en premier lieu les déterminants. Gross (1993) insiste pour sa part sur le caractère non compositionnel des phrases figées qui les définit, comme dans cet exemple :

[1] se boucher les yeux

Dans d'autres exemples comme

[2] ouvrir l'œil/ouvrir les yeux

on devra distinguer deux interprétations de l'expression, l'une littérale, qui en fait une construction verbale libre, l'autre métaphorique, qui en fait une construction verbale figée.

Au contraire, une expression comme

[3] avoir de bons yeux

ne manifeste qu'une interprétation régulière, sans figure : « avoir de bons yeux » signifie être doué d'une bonne vue.

Le bilan dressé par M. Gross à propos de la notion formelle se dessine ainsi : étant donnée une expression, on la dira figée quand deux de ses termes (au moins) sont indissociables, autrement dit figés l'un par rapport à l'autre.

1.2 Les formes des phrases figées contenant un « nom partie du corps »

On observe une grande variété formelle de ces phrases figées, mais elles partagent un point commun : cette notation se généralise, selon M. Gross, à des phrases dont le figement n'est pas complet. Dans les expressions retenues dans cette étude, le syntagme contenant un « nom partie du corps » au moins est figé, les autres syntagmes peuvent être figés également ou libres. Mais, d'autre part, il faut noter que

le figement du lien entre le verbe et le groupe nominal est variable. Ainsi, dans les phrases qui suivent, un des SN figés peut être le sujet ou un complément de rang 1, 2 ou 3, ou complément de nom, un éventuel second complément pouvant être libre ou figé :

[4] la moutarde lui monte au nez

[5] se boucher le nez

[6] avoir quelqu'un dans le nez

[7] avoir un coup dans le nez

[8] tirer les vers du nez

La détermination du figement peut faire difficulté. Ainsi, dans l'exemple

[9] mettre quelque chose sous les yeux de quelqu'un

les arguments pour analyser la construction du verbe comme figée sont de trois ordres :

1. la restriction du complément : elle n'est pas évidente puisqu'il existe

[10] mettre quelque chose sous le nez de quelqu'un

mais on ne peut faire varier la préposition :

[11] ? mettre quelque chose devant le nez/les yeux de quelqu'un.

2. le blocage du verbe qui refuse toute substitution

[12] *poser quelque chose sous le nez/les yeux de quelqu'un

3. l'opacité du sens : un locuteur non francophone de naissance aura du mal à comprendre l'expression

[13] avoir qqn dans le nez

On rencontre des difficultés pour décider de la position de certains exemples dans la classification des énoncés :

[14] lui mettre le nez dans sa crotte/ses ordures/son caca

C'est le caractère idiomatique – non compositionnel – des combinaisons verbe-groupe nominal et leur limitation à un faible nombre de substantifs qui constituent un argument en faveur du traitement de ces phrases comme des formes figées : on n'a pas affaire à un paradigme de compléments mais à des expressions synonymes. Dans ce cas, on note que toutes les phrases sont synonymes et que la liste des noms substituables est « fermée ». La variété des noms compléments n'induit pas de différences d'interprétation. On a donc affaire à une série figée qui correspond à une seule unité sémantique se présentant sous différentes variantes lexicales.

1.3 Problèmes d'identification des formes figées : constructions verbales vs suites adverbiales

L'expression

[15] coûter les yeux de la tête

pose un problème, celui des constituants figés qui ne sont pas des arguments, sujet ou objet du verbe, mais des modificateurs, des « adverbes » au sens de M. Gross (1986), qui marquent, par une structure figée, soit une qualité d'intensité, comme dans l'exemple ci-dessus, soit une expression de manière, comme dans l'exemple suivant :

[16] rire au nez de qqn

2 La traduction en italien des constructions verbales figées françaises contenant les noms de parties du corps *œil* et *nez*. Analyse d'un corpus

2.1 Le corpus français des constructions verbales figées contenant *œil* et *nez*

Pour établir le corpus français des constructions verbales figées contenant les deux noms « partie du corps » *œil* et *nez*, nous avons d'abord étudié le corpus mis au point à l'Université de Provence par André Valli et ses collaborateurs dans les années quatre-vingt-dix ; à partir de ce noyau, nous avons ensuite consulté les sources lexicographiques suivantes :

- un dictionnaire des expressions et locutions (Rey et Chantreau) ;
- trois dictionnaires de la langue française (*Petit Robert*, *Grand Robert*, TLFi).

Enfin, pour les cas douteux, nous avons consulté le Web, par le truchement de *Google*, et, si nécessaire, la compétence linguistique de quelques locuteurs natifs.

Nous avons ainsi construit un corpus français de 132 constructions verbales figées, dont 84 contenant le nom « partie du corps » *œil* et 48 contenant le nom « partie du corps » *nez*.

2.2 Classification des structures syntaxiques figées

Voici la classification des structures syntaxiques figées que nous avons repérées dans notre corpus : chaque structure est présentée d'abord sous forme de schéma formel¹ suivi du nombre d'occurrences entre parenthèses, et exemplifiée ensuite par une ou plusieurs constructions verbales figées (tableau 1).

¹ Dans un schéma formel, le symbole C indique, comme chez Maurice Gross, un argument figé ; si cet argument figé est représenté par un nom « partie du corps », le symbole C est suivi par le symbole Npc entre parenthèses — C (Npc).

| | ŒIL (84) | | NEZ (48) |
|----|--|----|--------------------------------------|
| 1. | No V C1 (Npc) (21) | 1. | No V C1 (Npc) (14) |
| | <i>ouvrir les yeux</i> | | <i>se casser le nez</i> |
| | <i>avoir de petits yeux/les yeux bouchés</i> | | <i>avoir bon nez/le nez fin</i> |
| | <i>avoir des yeux de lynx</i> | | |
| | <i>être tout yeux</i> | | |
| | <i>avoir un œil qui dit merde à l'autre</i> | | |
| | | | |
| 2. | No V Prép. C1 (Npc) (5) | 2. | No V Prép. C1 (Npc) (5) |
| | <i>tourner de l'œil</i> | | <i>piquer du nez</i> |
| | <i>sauter aux yeux de qqn</i> | | <i>passer sous le nez de qqn</i> |
| | | | |
| 3. | No V C1 (Npc) Prép. N2 (18) | 3. | No V C1 (Npc) Prép. N2 (9) |
| | <i>fermer les yeux sur qqch</i> | | <i>fouerrer son nez dans qqch</i> |
| | <i>faire les gros yeux à qqn</i> | | <i>faire un long nez à qqn</i> |
| | <i>faire des yeux de merlan frit à qqn</i> | | |
| | | | |
| 4. | No V N1 Prép. C2 (Npc) (13) | 4. | No V N1 Prép. C2 (Npc) (4) |
| | <i>regarder qqch d'un œil neuf</i> | | <i>avoir qqn dans le nez</i> |
| | <i>mettre qqch devant les yeux de qqn</i> | | |
| | | | |
| 5. | No V C1 (Npc) [Prép.] C2 (11) | 5. | No V C1 (Npc) [Prép.] C2 (5) |
| | <i>ouvrir les yeux à la lumière</i> | | <i>avoir le nez en l'air</i> |
| | <i>avoir les yeux partout</i> | | <i>fouerrer son nez partout</i> |
| | <i>ne plus avoir que ses yeux pour pleurer</i> | | |
| | | | |
| 6. | No V C1 Prép. C2 (Npc) (13) | 6. | No V C1 Prép. C2 (Npc) (5) |
| | <i>avoir le compas dans l'œil</i> | | <i>avoir un coup dans le nez</i> |
| | <i>jeter de la poudre aux yeux de qqn</i> | | <i>fermer la porte au nez de qqn</i> |
| | | | |
| 7. | No lui V Prép. C1 (Npc) (2) | 7. | No lui V Prép. C1 (Npc) (Ø) |
| | <i>lui sortir par les yeux</i> | | |
| | | | |
| 8. | Co lui V Prép. C1 (Npc), à N (Ø) | 8. | Co lui V Prép. C1 (Npc), à N (1) |
| | | | <i>la moutarde me monte au nez</i> |
| | | | |
| 9. | No V C1 (Npc) Prép. N2 Prép. N3 (1) | | No V C1 (Npc) Prép. N2 Prép. N3 (Ø) |
| | <i>ouvrir les yeux à qqn sur qqch</i> | | |

| | | | |
|-----|--|-----|--|
| 10. | No V C1 (Npc) Prép. C2 Prép. N3 (Ø) | 10. | No V C1 (Npc) Prép. C2 Prép. N3 (4) |
| | | | <i>mettre le nez dans sa merde à qqn</i> |
| 11. | No ne V pas Comparatif sur Adv (figé) dét. composé C (Npc) (Ø) | 11. | No ne V pas Comparatif sur Adv (figé) dét. composé C (Npc) (1) |
| | | | <i>ne pas voir plus loin que le bout de son nez</i> |

Tableau 1

D'après les données collectées, les deux structures syntaxiques les plus fréquentes dans notre corpus sont les suivantes : 1) la structure à deux arguments « No V C1 (Npc) », dont nous avons repéré 21 occurrences pour *œil* (25 %) et 14 occurrences pour *nez* (29 %) ; 2) la structure à trois arguments « No V N1 Prép N2 » dans ses quatre variantes :

- N0 V C1 (Npc) Prép. N2
- N0 V N1 Prép. C2 (Npc)
- N0 V C1 (Npc) [Prép.] C2
- N0 V C1 Prép. C2 (Npc)

dont nous avons repéré 55 occurrences pour *œil* (65,4 %) et 23 occurrences pour *nez* (47,85 %). La somme des constructions verbales figées françaises ayant recours à ces deux structures syntaxiques atteint, dans le cas de *œil*, 76 occurrences, soit 90,4 %, et, dans le cas de *nez*, 37 occurrences, soit 76,85 %, autrement dit l'immense majorité des constructions verbales figées dans les deux cas.

3 Le corpus bilingue français-italien

3.1 Méthodologie

Pour établir la liste des constructions verbales figées italiennes correspondant aux constructions verbales figées françaises faisant partie de notre corpus, nous avons d'abord consulté les sources lexicographiques suivantes :

- trois dictionnaires bilingues (Boch, DIE, Garzanti) ;
- trois dictionnaires monolingues (DISC, GRADIT, Treccani) ;
- un dictionnaire des locutions italiennes (Quartu).

Ensuite, pour les cas douteux, nous avons consulté le Web, par le truchement de *Google*, et, si nécessaire, la compétence linguistique de quelques locuteurs natifs.

Nous avons enfin adopté la solution la plus proche, du point de vue sémantique et syntaxique, de la construction verbale figée française.

3.2 Deux problèmes d'ordre sémantique

Dans ce travail d'identification des constructions verbales figées italiennes, nous avons surtout rencontré deux problèmes d'ordre sémantique : 1) la polysémie des constructions verbales figées françaises ; 2) la polysémie de certaines constructions verbales figées italiennes, homologues d'une construction verbale figée française. Examinons en détail ces deux problèmes qui représentent les deux faces de la même médaille.

Certaines constructions verbales figées françaises ont plusieurs sens, selon le contexte où elles sont employées. Dans quelques cas, lorsque nous n'avons pas repéré en italien une construction verbale figée couvrant la même pluralité de sens, nous avons été obligés de séparer les différents sens de la construction figée française et de proposer des traductions différentes. C'est le cas de la construction verbale figée française

[17] avoir le nez sur qqch → avoir le nez dessus

qui signifie aussi bien « ne pas voir ce qui est trop près » que « rester plongé dans une occupation, sans se laisser distraire ». Or, la construction verbale correspondante en italien – *avere il naso su qsa* → *averci il naso sopra* – couvre le premier sens mais non pas le deuxième, pour lequel nous n'avons pas repéré de construction figée. D'où la nécessité de séparer les deux sens de la construction figée française et de proposer deux traductions différentes.

Dans d'autres cas, la correspondance sémantique entre une construction figée française et une construction figée italienne, syntaxiquement homologues, n'est que partielle, la construction italienne exprimant un sens plus large que la construction correspondante française. C'est le cas, par exemple, de la construction italienne

[18] *fare l'occhiolino*

que les dictionnaires bilingues proposent pour traduire l'expression française *faire de l'œil*, qui signifie « adresser des regards amoureux » (Rey et Chantreau 1997) dans l'intention de séduire. Or *fare l'occhiolino* signifie plus généralement « cligner de l'œil » soit en signe de connivence – *in segno d'intesa* (Treccani) – soit dans une intention grivoise – *con intenzione maliziosa* (Treccani). L'expression italienne n'aura donc le sens de l'expression française que dans certains contextes.

Étant donné que notre étude sur la traduction en italien des constructions verbales figées françaises contenant un nom « partie du corps » n'en est encore qu'à ses débuts, nous avons décidé d'accepter provisoirement ces solutions, quoiqu'imparfaites.

3.3 Critères de correspondance

Les différents types de relations syntaxiques entre constructions verbales figées françaises et italiennes ont été répartis en cinq classes : 1) homologie totale ; 2) homologie partielle ; 3) expression figée italienne contenant un autre nom « partie du

corps » ; 4) expression figée italienne ne contenant pas de nom « partie du corps » ; 5) absence d'expression figée en italien.

1. L'« homologie totale » (HT) correspond à « une identité totale sur le plan syntaxique » entre la construction figée française et la construction figée italienne, « comprenant la même partie du corps et développant les mêmes significations, littéraire et métaphorique » (Valli, Villagenes Serra 1998 :187). Voici quelques exemples :

[19] ouvrir les yeux → *aprire gli occhi*

[20] parler aux yeux → *parlare agli occhi*

[21] faire les yeux doux à qqn → *fare gli occhi dolci a qno*

Dans certains cas, l'absence de correspondance entre la structure syntaxique de la construction figée française et celle de la construction figée italienne tient à une règle grammaticale de l'italien : c'est le cas, par exemple, de la traduction en italien de l'article partitif français ou de certaines prépositions françaises. Bien qu'en italien l'article partitif existe, son emploi est plus restreint, l'italien préférant souvent avoir recours à un déterminant \emptyset :

[22] avoir des yeux de lynx → *avere occhi di lince*

Nous avons classé ce type de relation, où l'absence de correspondance au niveau du déterminant est imposée par la grammaire, parmi les homologies totales². De même pour certains compléments prépositionnels : bien que le système prépositionnel de l'italien soit semblable au système français, il existe tout de même des différences d'ordre grammatical. Nous avons repéré trois cas dans notre corpus :

- i. le complément « à N » à valeur locative : même si la préposition italienne *a* peut avoir aussi un sens locatif, dans plusieurs cas un complément locatif français « à N » se traduit en italien par « in N » :

[23] jeter de la poudre aux yeux de qqn → *gettare polvere negli occhi di qno*³

- ii. le complément « de N_{pc} » à valeur de moyen/manière : pour traduire ce type de complément on n'a pas recours en italien à l'une des prépositions correspondantes, *di* ou *da*, mais à la préposition *con*, comme dans l'exemple suivant :

[24] manger qqn des yeux → *mangiare qno con gli occhi*⁴

- iii. le locatif de passage « par N » : comme il n'existe pas en italien une préposition correspondant à *par*, on traduit ce type de locatif soit par la préposition *attraverso* soit par la préposition *da*, comme dans l'exemple suivant :

2 Nous avons repéré dans notre corpus cinq exemples de traduction d'un article partitif français par un déterminant \emptyset et un seul exemple de traduction d'un article indéfini français par un déterminant \emptyset . Tous ces exemples ont été classés parmi les homologies totales.

3 Nous avons repéré un seul cas de ce type de relation dans notre corpus.

4 Nous avons repéré quatre cas de ce type de relation dans notre corpus.

[25] sortir par les yeux à qqn → *uscire dagli occhi a qno*⁵

Nous avons classé ces trois types de relations prépositionnelles imposées par la grammaire parmi les homologues totales.

2. L'« homologues partielle » (HP) correspond à une identité de structure syntaxique, à un élément près. Cet élément qui diffère peut être soit le nombre du Npc (se rincer l'œil → *lustrarsi gli occhi*), soit le déterminant du Npc (ne pas fermer l'œil → *non chiudere occhio*), soit le modificateur du Npc (avoir l'œil aux aguets → *avere l'occhio vigile*), soit une préposition (ne pas en croire ses yeux → *non credere ai propri occhi*), soit un deuxième complément figé du verbe (ne rien voir qu'avec les yeux de qqn → *vedere tutto con gli occhi di qno*), soit le verbe lui-même (boire qqn des yeux → *mangiare qno con gli occhi*).

Nous avons aussi considéré comme des homologues partielles les cas où la construction figée italienne, tout en gardant le même Npc, présente plusieurs variantes au niveau des autres éléments de la structure, comme dans les exemples suivants :

- [26] crever les yeux → *saltare agli occhi* [verbe + prép.]
 [27] avoir qqn à l'œil → *tenere qno d'occhio* [verbe + prép. + dét.]
 [28] avoir l'œil à tout → *tenere tutto d'occhio* [tout sauf Npc]

3. Dans quelques cas, la traduction en italien d'une construction figée française est une construction figée contenant un autre nom « partie du corps », comme dans les exemples suivants :

- [29] faire un long nez à qqn → *fare il muso a qno*
 [30] claquer la porte au nez de qqn → *sbattere la porta in faccia a qno*

4. Dans d'autres cas, la traduction en italien d'une construction figée française est une construction figée ne contenant aucun nom « partie du corps », comme dans l'exemple suivant :

- [31] ne pas avoir les yeux en face des trous → *non vederci dal sonno, dalla stanchezza*

5. Le dernier type de relation que nous avons repéré dans notre corpus correspond au cas où il n'existe pas en italien une construction verbale figée pour traduire une certaine construction figée française : dans ce cas-là, il faudra avoir recours à la définition sémantique de la construction française.

- [32] avoir des yeux au bout des doigts → [faire avec habileté des ouvrages de la main]

⁵ Nous avons repéré un seul cas de ce type de relation dans notre corpus.

3.4 Tableau des correspondances entre constructions verbales figées françaises et italiennes

Les données définitives concernant les différents types de relations syntaxiques repérés dans notre corpus entre constructions verbales figées françaises et italiennes contenant les noms « partie du corps » *œil/occhio* et *nez/naso* ont été regroupées dans le tableau 2 ci-dessous :

| RÉSULTATS | œil = 84 | | nez = 48 | |
|---------------------------|----------|------|----------|-------|
| HT | 37 | 44 | 12 | 25 |
| HP nombre | 4 | 4,75 | | |
| HP déterminant | 2 | 2,4 | 8 | 16,65 |
| HP modificateur | 8 | 9,5 | 1 | 2,1 |
| HP préposition | 2 | 2,4 | | |
| HP (C1 ou C2) | 3 | 3,6 | | |
| HP verbe | 4 | 4,75 | 1 | 2,1 |
| HP combinaisons | 13 | 15,5 | 3 | 6,25 |
| total HP | 36 | 42,9 | 13 | 27,1 |
| total HT + HP | | 86,9 | | 52,1 |
| Autre Npc | 1 | 1,2 | 15 | 31,25 |
| Autre constr. verb. figée | 6 | 7,15 | 2 | 4,15 |
| Absence d'expr. figée | 4 | 4,75 | 6 | 12,5 |

Tableau 2

Les données affichées dans le tableau ci-dessus montrent d'une manière claire que la traduction en italien des constructions verbales figées françaises contenant le nom « partie du corps » *œil* est beaucoup plus stable, et donc plus facile à prévoir, que la traduction des constructions verbales figées françaises contenant le nom « partie du corps » *nez*.

En particulier, 44 % des constructions françaises contenant *œil* sont dans une relation d'homologie totale (HT) avec les constructions italiennes correspondantes, alors que seulement 25 % des constructions françaises contenant *nez* sont dans cette relation avec les constructions italiennes correspondantes.

Si l'on considère globalement les différents types d'homologie partielle (HP), on obtient un résultat similaire: 42,9 % des constructions françaises contenant *œil* sont dans une relation d'homologie partielle avec les constructions italiennes correspondantes, alors que seulement 27,1 % des constructions françaises contenant *nez* sont dans cette relation avec les constructions italiennes correspondantes.

La somme de ces données montre que 86,9 % des constructions françaises contenant *œil* sont dans une relation d'homologie totale ou partielle avec les constructions italiennes correspondantes : la métaphore ontologique de l'œil se retrouve donc de façon très stable dans les deux langues en présence.

Au contraire, la somme des données concernant les constructions françaises contenant *nez* montre que seulement 52,1 % de ces constructions sont dans une relation d'homologie totale ou partielle avec les constructions italiennes correspondantes : en effet, pour plus de 30 % des constructions figées françaises contenant ce nom « partie du corps » (31,25 %), *nez* est remplacé dans la construction correspondante italienne par un autre nom « partie du corps » (dans la plupart des cas, *faccia, muso, occhi*).

Sur la base de ces résultats, notre hypothèse est que la traduction en italien des constructions figées françaises contenant *œil* est stable parce que cette partie du corps est considérée comme un organe de sens dans les deux langues, alors que la traduction en italien des constructions figées françaises contenant *nez* est assez instable parce que cette partie du corps est souvent considérée en français comme un élément du visage, ce qui permet de la remplacer en italien par un autre élément du visage ou par un nom désignant le visage lui-même (il apparaît alors une relation de métonymie « partie/tout »).

Bibliographie

A. Essais critiques

- Bercker (F.), 1992 : « Alouette, gentille alouette... Parties du corps et constructions figées », dans Perrin (L.), dir., *Approche énonciative de l'énoncé complexe*, Louvain, Peters-Bibliothèque de l'information grammaticale, p. 119-143.
- Blanche-Benveniste (C.) et Valli (A.), 1997 : « L'expérience Eurom4 : comment négocier les difficultés », dans *L'intercompréhension : le cas des langues romanes. Le français dans le monde*, N° spécial coordonné par Claire Blanche-Benveniste et André Valli, janvier.
- Danlos (L.), 1988 : « Les expressions figées construites avec le verbe être Prep », dans *Langages*, n° 90, p. 23-39.
- Gross (M.), 1983 : « Les déterminants dans les expressions figées », dans *Langages*, n° 79, p. 89-117.
- Gross (M.), 1988 : « Les limites de la phrase figée », dans *Langages*, n° 90, p. 7-23.
- Gross (M.), 1993 : « Les phrases figées en français », dans *L'information grammaticale*, n° 59, p. 36-42.
- Lakoff (G.) et Johnson (M.), 1980 : *Metaphors we live by*, Chicago, The University of Chicago Press (trad. fr., *Les métaphores dans la vie quotidienne*, Paris, Éditions de Minuit, 1985).
- Lakoff (G.), 1987 : *Women, Fire, and Dangerous Things*, Chicago, The University of Chicago Press.
- Martins-Baltar (M.), 1977 : « Repères dans les recherches actuelles sur la locution », dans Martins-Baltar (M.), éd., *La locution entre langue et usages*, Fontenay/Saint-Cloud, ENS Éditions, p. 19-52.

Valli (A.) et Villagenes Serra (E.), 1998 : « Locutions figées comprenant un nom « partie du corps » en espagnol et en français », dans Mejri (S.) *et al.*, éd., *Le figement lexical, Rencontres linguistiques méditerranéennes*, Tunis, les 17-18 et 19 septembre 1998, p. 177-206.

B. Dictionnaires

1 Langue française

Rey (A.) et Chantreau (S.), 1997 : *Dictionnaire des expressions et locutions*, Paris, Le Robert.

Le Grand Robert. *Dictionnaire alphabétique et analogique de la langue française*, 2005, version électronique 2.0 sur cd-rom, Paris, Dictionnaires Le Robert.

Le Nouveau Petit Robert. *Dictionnaire alphabétique et analogique de la langue française*, 1997, version électronique sur cd-rom, Paris, Dictionnaires Le Robert.

T.L.F.I., 1994 : *Le Trésor de la Langue Française Informatisé*, Paris, CNRS Éditions, <http://atilf.atilf.fr/tlf.htm>

2 Bilingues français-italien

Boch (R.), dir., 2000 : *Il Boch quarta edizione in cd-rom. Dizionario francese-italiano italiano-francese*, Bologna, Zanichelli.

DIE, 2002 : *Dizionario Francese Italiano — Italiano Francese*, version électronique sur cd-rom, Milano, Paravia Bruno Mondadori.

Garzanti, 2003 : *Dizionario Interattivo Garzanti francese-italiano italiano-francese*, version électronique sur cd-rom, Milano, Garzanti.

3 Langue italienne

Quartu (B.M.), 2000 : *Dizionario dei modi di dire della lingua italiana*, Milano, Rizzoli Bur.

Sabatini (F.) et Coletti (V.), dir., 2005³ : *DISC. Il Sabatini Coletti. Dizionario della Lingua Italiana*, Milano, Rizzoli Larousse.

De Mauro (T.), dir., 2000 : *GRADIT. GRAnde Dizionario ITaliano dell'uso*, version électronique sur cd-rom, Torino, Utet.

Treccani, 1997 : *Il Vocabolario Treccani*, version électronique sur cd-rom, Roma, Istituto della Enciclopedia Italiana.

Van Dai Vu,
Université de Hanoï,
Vietnam,
daiphap@hanu.edu.vn.

Métaphore conceptuelle en traduction juridique et équivalence

Résumé

La traduction juridique est devenue essentielle aujourd'hui. La multiplicité des types textuels en droit et les différences entre systèmes juridiques sont un défi permanent pour le traducteur. Mais ce n'est pas là le seul élément de complexité dans ce type de traduction. Les concepts juridiques eux-mêmes diffèrent d'une langue à l'autre. Aussi est-il difficile de trouver l'équivalent adéquat pour les termes du droit. De même, le raisonnement juridique qui sous-tend les textes à traduire est variable selon les textes. De plus, les législateurs expriment leurs idées en ayant souvent recours à des expressions métaphoriques. Face à ces difficultés, il est donc important pour le traducteur de comprendre et maîtriser la logique du droit dans ses différentes langues de travail et de pouvoir passer du système conceptuel auquel recourt la langue source à celui de la langue cible. En d'autres termes, pour le domaine juridique, le traducteur devrait savoir s'il y a ou s'il n'y a pas correspondance entre les métaphores conceptuelles véhiculées par les langues en contact, et quelles en sont les conséquences sur ses méthodes de traduction tant en ce qui concerne le lexique que la phraséologie. La réponse à ces questions lui permet de s'affranchir de la recherche étroite de la correspondance lexicale pour prendre en considération les connaissances encyclopédiques sous l'angle de leur organisation conceptuelle, de produire un texte équivalent en langue d'arrivée sans trahir l'esprit des lois originales. Pour illustrer cette proposition, dans sa communication l'auteur prend appui sur plusieurs exemples issus de deux systèmes juridiques français et vietnamien et procède à une analyse des métaphores conceptuelles afin de faire ressortir la diversité des situations et des approches de la traduction juridique.

Mots-clés : traduction juridique ; concepts juridiques ; correspondance ; équivalence ; expressions métaphoriques ; métaphore conceptuelle.

Introduction

Ces dernières années, les études juridiques constituent un terrain d'études particulièrement intéressant au Vietnam. Il s'agit d'un domaine dynamique qui s'est considérablement développé. D'une part, le pays accélère son processus d'intégration dans le monde avec la construction d'une économie de marché et l'adhésion à plusieurs organisations mondiales, dont l'OMC. D'autre part, les investisseurs

étrangers y trouvent un très bon environnement pour leurs affaires. Dans ce contexte, la connaissance des lois nationales ainsi qu'internationales est aussi fondamentale pour le Vietnam que pour ses partenaires étrangers. La traduction juridique est alors devenue essentielle pour le bon fonctionnement des institutions et des sociétés. Mais on constate que ce type de traduction provoque pour le traducteur de nombreuses difficultés. D'abord, la multiplicité des types textuels en droit et les différences entre systèmes juridiques sont un défi permanent pour le traducteur. Mais ce n'est pas là le seul élément de complexité dans ce type de traduction. Les concepts juridiques eux-mêmes diffèrent d'une langue à l'autre. Aussi est-il difficile de trouver l'équivalent adéquat pour les termes du droit. De même, le raisonnement juridique qui sous-tend les textes à traduire est variable selon les textes. De plus, les législateurs expriment leurs idées en ayant souvent recours à des expressions métaphoriques, qui, selon Lakoff (1996) et ses collaborateurs, sont l'expression de surface de « métaphores conceptuelles » structurant un domaine. La métaphore conceptuelle est définie comme l'« établissement de correspondance entre un domaine source et un domaine cible ». Ce constat nous a amené dans cet article à considérer à partir d'un corpus, à savoir la loi vietnamienne sur la famille et le mariage et sa traduction française, comment sont reconstruits dans la langue cible les concepts métaphorisés du texte source tant en ce qui concerne le lexique que la phraséologie. La mise en évidence de ces points permet de s'affranchir de la recherche étroite de la correspondance lexicale pour prendre en considération les connaissances encyclopédiques sous l'angle de leur organisation conceptuelle, de produire un texte équivalent en langue d'arrivée sans trahir l'esprit des lois originales.

1 Choix du texte de départ

Le document retenu pour notre étude est la loi sur la famille et le mariage N° 22/2000/QH10, promulguée par l'Assemblée nationale du Vietnam le 9 juin 2000. La traduction française de ce texte est réalisée par la Maison du droit vietnamo-française (M.D.V.F.), qui créée en 1993 et entièrement consacrée aux questions juridiques, est « un lieu d'échanges entre juristes des deux pays, un espace de réflexion sur les évolutions juridiques accompagnant l'ouverture politique et économique du Vietnam, et qui est l'outil principal de la coopération juridique franco-vietnamienne ».

Ce texte a été choisi pour les raisons suivantes :

- c'est un texte officiel du domaine juridique ;
- il comporte des expressions métaphoriques intéressantes ;
- sa traduction française est assurée par la Maison du droit vietnamo-française, un établissement prestigieux dans la traduction juridique.

Il n'est pas possible de reproduire la totalité du texte de loi dans la limite d'un article, ni de faire l'inventaire de toutes les expressions métaphoriques du texte. Pour analyser les problèmes que nous avons soulevés, nous en avons sélectionné

quatre d'entre elles. L'énoncé original porte un chiffre et la traduction du M.D.V.F., est signalée par un chiffre prime, ce qui nous donne :

- [1] *Gia đình là tế bào của xã hội*
 Famille **être cellule** de société

[1'] La famille est une cellule de la société

- [2] *Gia đình là cái nôi nuôi dưỡng con người, là môi trường quan trọng hình thành và giáo dục nhân cách, góp phần vào sự nghiệp xây dựng và bảo vệ Tổ quốc.*

famille **être berceau** entretenir homme être environnement important donner naissance éduquer personnalité contribuer œuvre construire et édifier patrie

[2'] La famille est le berceau dans lequel chacun est élevé, le lieu de formation et de développement de la personnalité, et contribue ainsi à l'œuvre de construction et de défense de la nation.

- [3] *Anh chị em có bổn phận thương yêu chăm sóc giúp đỡ lẫn nhau; có nghĩa vụ và có quyền đùm bọc nuôi dưỡng nhau trong trường hợp không có cha mẹ hoặc cha mẹ không có điều kiện trông nom.*

Frères sœurs avoir obligation s'aimer s'entraider, avoir devoir et droit **se protéger** s'entretenir en cas de ne pas avoir parents ou parents ne pas avoir possibilité de prendre soins

[3'] Les frères et sœurs ont, l'un à l'égard de l'autre, le devoir d'affection, d'entretien et d'assistance; ils ont le droit et l'obligation de s'entraider et de s'entretenir mutuellement si leurs parents sont décédés ou ne peuvent assurer leur garde, leur entretien et leur éducation.

- [4] *Bố dượng, mẹ kế có nghĩa vụ và quyền trông nom, nuôi dưỡng, chăm sóc giáo dục con riêng cùng sống chung với mình theo quy định tại các điều 34,36,37 của Luật này.*

beau-père, belle-mère avoir obligation et droit de prendre soin, entretenir, éduquer **enfant propre** ensemble habiter avec selon règlement article 34,36, 37 de loi ci

[4'] Le beau-père ou la belle-mère a le droit et le devoir de garde, d'entretien et d'éducation à l'égard de l'enfant du premier lit de son conjoint et vivant dans la seconde famille de ce dernier, conformément aux dispositions des articles 34,36 et 37 de la présente Loi.

2 Métaphore conceptuelle

L'observation de notre petit corpus permet de dégager quelques concepts métaphoriques : *famille, enfant, frère et sœur*. Pour analyser le mécanisme de la métaphorisation conceptuelle, nous nous référons au modèle de métaphore conceptuelle développée par Lakoff, qui a mis en évidence l'importance des métaphores

conceptuelles dans l'usage quotidien de la langue, et qui a défini la métaphore comme un phénomène de pensée, la projection d'un cadre conceptuel sur un autre.

L'exemple de l'emploi du terme *incubateur* est une bonne illustration de la pensée de Lakoff. En effet, le dictionnaire *Larousse* explique le sens de ce mot comme suit :

- Appareil servant à l'incubation artificielle des œufs de poule, d'oie, etc. Synonyme de *couveuse*.
- Économie : Structure créée par de grands groupes, réunissant des start-ups dont ils financent le lancement ou dans lesquelles ils prennent des participations, pour en favoriser la croissance.

Le sens métaphorique du terme correspond à ce qu'on appelle *incubateur d'entreprise*, qui est défini par Wikipédia comme suit :

« Un incubateur d'entreprises est une structure d'accompagnement de projets de création d'entreprises. L'incubateur peut apporter un appui en termes d'hébergement, de conseil et de financement, lors des premières étapes de la vie de l'entreprise. À la différence d'une pépinière ou d'un hôtel d'entreprises, un incubateur s'adresse à des sociétés très jeunes ou encore en création, et leur propose un ensemble de services adaptés. Les incubateurs peuvent se différencier entre eux par les services qu'ils proposent, leur caractère lucratif ou non, ou encore le type de projets qu'ils ciblent. »

3 Représentation conceptuelle des métaphores dans le texte source : la chosification

La chosification fait partie intégrante du système de métaphorisation, dont elle est une catégorie. Dans nos exemples [1] et [2], les concepts « famille », « enfant », « frères et sœurs » sont compris comme s'ils étaient des choses : la famille est perçue comme si elle était une unité structurale fonctionnelle et reproductrice constituant tout ou partie d'un être vivant, une cellule. Elle est aussi représentée comme un lit particulier (un berceau) dans lequel dorment les nourrissons et les petits enfants jusque deux à trois ans, ou plus s'il est de grande taille. Il s'agit d'une projection d'un domaine physique (berceau, qu'on peut balancer pour *bercer*) sur un domaine immatériel (famille).

L'exemple [3] est particulièrement intéressant du point de vue de la métaphorisation. En effet, l'emploi du verbe composé *đùm bọc* qui figure en gras dans l'énoncé original a donné lieu à une chosification :

Anh chị em có bổn phận thương yêu chăm sóc giúp đỡ lẫn nhau; có nghĩa vụ và có quyền đùm bọc nuôi dưỡng nhau .

Frères sœurs avoir obligation s'aimer s'entraider, avoir devoir et droit **se protéger**

Dans la langue vietnamienne, *đùm* et *bọc* sont des verbes simples et quasi synonymes qui signifient respectivement *envelopper*, *couvrir*. Ces unités simples peuvent se combiner pour former un verbe composé et signifier alors *se protéger* et *s'aider*. Le mécanisme de métaphorisation est le suivant : les frères et sœurs privés de parents ou de tuteurs, donc orphelins et en situation difficile, sont représentés comme des objets fragiles nécessitant ainsi une protection et un soin particuliers.

L'exemple [4] comporte un concept faisant partie de la loi matrimoniale et familiale : il s'agit de *con riêng* (l'enfant propre au mari ou à la femme, c'est-à-dire, celui ou celle du premier mariage). Cet exemple illustre bien une situation de non-correspondance entre les langues dans le domaine de la métaphore conceptuelle que nous analysons dans le paragraphe suivant.

4 Représentation conceptuelle des métaphores dans la traduction française

Comme nous l'avons signalé, la traduction française de la loi qui fait l'objet de la présente étude est assurée par la Maison du droit vietnamo-française. Il ne s'agit pas pour nous de procéder à une évaluation de la traduction, mais de voir s'il y a un recoupement ou un écart conceptuel entre la formulation de départ et la formulation d'arrivée. Nous nous abstenons donc de tout commentaire sur la qualité de la traduction. De plus, faute de traductions réalisées par d'autres traducteurs, nous ne pouvons pas non plus faire de remarques sur d'éventuelles différences entre des choix de traduction. Il nous intéresse par contre de voir s'il y a équivalence de concepts à la suite d'un transfert des expressions métaphoriques, de vérifier si le lien conceptuel entre le texte source et le texte cible est respecté.

Dans la traduction de l'énoncé [1], la formulation française correspond bien à l'expression originale. C'est parce que le traducteur a opté pour une traduction littérale. Cette solution est-elle acceptable ? La réponse est positive puisqu'il y a un bon recoupement conceptuel entre l'énoncé original et sa traduction française. En effet, on trouve en français des expressions telles que « la famille, cellule de la société »¹, qui veut dire que la famille est un élément constitutif d'un organisme vivant qu'est la société : « la famille est la cellule de base de la communauté humaine » ; « la famille est la cellule vitale de la société »². Ainsi, la conceptualisation de la notion de famille est la même pour le texte de départ et le texte d'arrivée : la famille est perçue comme un élément constitutif, une unité morphologique et fonctionnelle constitutive de tout être vivant.

1 Voir le *Petit Robert* 1993, p. 326, entrée *Famille*

2 <http://www.eglise.catholique.fr/eglise-et-société>

Considérons maintenant le cas [2] où la famille est conceptualisée comme un « berceau » dans le texte de départ. Dans la langue courante, le berceau est défini comme suit³ :

« Petit lit où l'on couche les nourrissons et qui est disposé pour qu'on puisse le balancer, le bercer, aisément : **Berceau** d'osier Mettre un enfant dans son **berceau**. Un enfant qui est encore au **berceau**. »

Dans son sens figuré, ce mot renvoie à un lieu de naissance, un commencement, un début :

« Florence fut le **berceau** de la peinture moderne. La Saxe fut le **berceau** du luthéranisme. La plupart des historiens regardent l'Inde comme le **berceau** de la civilisation. »

« La famille, **berceau** de vocations ! »

Notre connaissance de la structure du domaine source (famille) et le contexte linguistique nous fait rejeter le sens propre de l'énoncé. C'est donc le sens figuré d'être le premier lieu de formation et de développement de la personnalité qui est spontanément retenu. La saisie de ce concept permet d'adopter une solution de traduction vers le français : *La famille est le berceau dans lequel chacun est élevé, le lieu de formation et de développement de la personnalité, et contribue ainsi à l'œuvre de construction et de défense de la nation.*

La comparaison de l'énoncé original et sa traduction française montre que la traduction sémantique qui appelle des transformations lexico-syntaxiques peut dans certains cas assurer le lien conceptuel entre les langues, et ce, à condition que les peuples qui les parlent aient la même manière de penser, le même mode de conceptualisation. C'est le cas ici, d'où l'acceptabilité de la traduction proposée du point de vue de l'équivalence conceptuelle.

Examinons la représentation conceptuelle dans la métaphore *đùm bọc* (*envelopper, couvrir*) de l'exemple [3]. Cette représentation peut être illustrée par le biais d'une projection domaniale au moyen du schéma suivant :

envelopper et couvrir une chose pour la protéger

frères et sœurs s'envelopper, se couvrir pour se protéger de tout risque et danger

Dans la traduction de la M.D.V.F. le sens figuré du verbe composé *đùm bọc* (*envelopper, couvrir*) a été rendu par *s'entraider et s'entretenir mutuellement*. On constate facilement que la version française ne peut pas maintenir l'image métaphorique du texte d'origine, à savoir : les frères et sœurs sont perçus comme quelque chose de fragile en l'absence des parents ou de tuteur. Comme l'ont remarqué les fondateurs de la théorie de la métaphore conceptuelle, celle-ci est un phénomène de pensée, de conceptualisation. Dans notre cas il y a une variance de pensée, de mode de conceptualisation qui aboutit à la non-correspondance d'expression linguistique.

3 <http://fr.wiktionary.org/wiki/berceau>

Si le vietnamien emploie une métaphore qui consiste en une chosification, le français n'a pas recours à ce moyen. Ayant observé ce phénomène, le traducteur n'a pas procédé à un transcodage linguistique en risquant de provoquer un problème de compréhension mais a choisi la traduction par équivalence, qui exige une « récréation contextuelle » selon la théorie interprétative de la traduction de l'école de Paris (Séleskovitch 1986, Delisle 1980). Mais on remarque que cette récréation ne doit pas détruire le lien conceptuel entre l'expression métaphorique du texte source et l'expression équivalente du texte cible.

Nous avons annoncé ci-dessus que la traduction de l'énoncé [4] est une bonne illustration de la non-correspondance entre les langues dans le domaine de la métaphore conceptuelle. Pour le concept « l'enfant du mari ou de la femme de son premier mariage », le vietnamien emploie le terme *con riêng* (littéralement : *propre enfant*). Par contre dans la langue française, à côté de l'expression *enfant du premier mariage*, nous trouvons *enfant du premier lit*, concept qui n'existe pas en vietnamien. Dans cette situation, pour arriver à une solution convenable en français, le traducteur peut adopter plusieurs approches. Soit il se fie à une traduction réflexe déclenchée par une interférence linguistique, en disant :

« Le beau-père ou la belle-mère a le droit et le devoir de garde, d'entretien et d'éducation à l'égard **de l'enfant propre à son conjoint** et vivant dans la seconde famille de ce dernier, conformément aux dispositions des articles 34, 36 et 37 de la présente Loi. »

Soit, il peut expliquer le concept original dans la langue d'arrivée par une expression équivalente au niveau linguistique et conceptuel en adoptant la solution : [...] ***de l'enfant du premier mariage de son conjoint et vivant dans la seconde famille.***

Soit, le traducteur peut avoir recours à l'expression métaphorique en français [...] ***de l'enfant du premier lit de son conjoint et vivant dans la seconde famille.***

Ces trois attitudes sont toutes repérées chez différentes catégories de traducteurs. La première est celle qu'adoptent souvent les apprenants de la traduction, en raison de leur insuffisance de compétence et d'expérience. C'est ce que nous avons observé au cours de notre travail d'enseignant de la traduction générale et spécialisée dans notre établissement. Par contre, les traducteurs professionnels n'acceptent pas facilement le transcodage linguistique. En effet, il ressort des entretiens que nous avons eus avec eux⁴, qu'ils se demandent toujours si l'expression retenue exprime bien le concept en question dans les langues, si elle convient aux habitudes langagières de la communauté d'accueil. Après une recherche documentaire avec des outils informatiques, et la demande d'avis des experts du domaine, s'il appert que ce n'est pas le cas, ils optent pour une autre expression plus idiomatique en essayant de maintenir intact le lien conceptuel entre le texte de départ et

4 Il s'agit d'entretiens avec des traducteurs professionnels de la Maison du droit vietnamo-française, de l'Agence vietnamienne d'information, et des offices de notariat à Hanoi.

le texte d'arrivée. Cela aboutit à une traduction par équivalence (*l'enfant du premier mariage de son conjoint*).

La solution qui consiste à utiliser la métaphore dans la langue d'arrivée *l'enfant du premier lit* est intéressante. D'abord, il n'est pas certain que tous les traducteurs connaissent cette métaphore. Par conséquent, la traduction par équivalence est appliquée dans la majorité des cas. Pour ceux qui se sont spécialisés dans les études juridiques, il y a de fortes chances qu'ils connaissent cette expression métaphorique mais ils ont à choisir entre l'expression métaphorique et non métaphorique. Ils ne sont pas forcés par ailleurs d'employer la tournure métaphorique car l'expression de la langue courante *l'enfant du premier mariage* est parfaitement claire et intelligible. La diversité des formulations retenues dans la langue cible pour traduire un énoncé original pose ainsi le problème de l'identification des représentations cognitives évoquées par l'expression métaphorique dans le texte de départ et montre la multiplicité et l'originalité des cas de recréation contextuelle. Les traducteurs de la M.D.V.F. ont choisi une expression métaphorique française pour réexprimer l'énoncé vietnamien. Mais ce n'est pas une solution unique. Le concept dans la langue de départ peut être rendu par une autre solution qui est aussi fidèle du point de vue du recoupement conceptuel.

Conclusion

Nous avons soulevé dans cet article des problèmes concernant le recoupement conceptuel entre des énoncés originaux et leur traduction, avec l'analyse de quatre métaphores, qui représentent chacune une image conceptuelle particulière. Nous avons aussi montré quelle traduction est plus facile à réaliser qu'une autre et quelles démarches le traducteur peut suivre pour trouver une équivalence conceptuelle. Il s'agit en première phase du processus traductionnel consistant à identifier le concept métaphorisé en mobilisant des connaissances encyclopédiques et thématiques dans le texte de départ. Ensuite le traducteur devra vérifier s'il y a correspondance ou non entre les concepts en question dans les langues et les expressions linguistiques qui les expriment en vue de trancher en faveur d'une solution de traduction. En tout cas, celle-ci doit prendre en compte les modes de conceptualisation et la correspondance ou non des concepts dans les langues de travail. En phase de reformulation, le traducteur pourra décider soit de reproduire dans la langue source les mêmes images conceptuelles, c'est-à-dire, transposer le même mode de conceptualisation, soit de rechercher une équivalence conceptuelle, en variant le mode de conceptualisation. C'est cette traduction, qui est plus difficile à réaliser mais qui est plus riche et plus nuancée, qui constitue un défi pour le traducteur dans un domaine de spécialité.

Bibliographie

- Delisle (J.), 1980 : *L'Analyse du discours comme méthode de traduction. Initiation à la traduction française de textes pragmatiques anglais. Théorie et pratique*, Ottawa, Presses de l'Université d'Ottawa.
- Lakoff (G.), 1996 : « Les universaux de la pensée métaphorique : la question de la variation dans l'expression linguistique », conférence tenue à l'ENA, Paris, lors du colloque *Diversité des langues et représentation cognitive*, Paris.
- Lederer (M.), 1981 : *La traduction simultanée, expérience et théorie*. Paris, Minard, Lettres Modernes.
- Rydning (A.F.), 2002 : « Concept métaphorique et expression métaphorique dans une perspective cognitiviste », dans *Romansk Forum XV Skandinaviske romanistkongress*, n° 16 – 2002/2 Oslo, p. 12.-17.
- Rydning (A.F.), 2002 : « Pénétrer la boîte noire du traducteur », dans *Linguistica Antverpiensia*, Amsterdam, p. 273-285.
- Seleskovich (D.) et Lederer (M.) 1986 : *Interpréter pour traduire*, Paris, Didier.
- Vandaele (S.), 2000 : « Métaphore conceptuelle et traduction biomédicale », dans *La traduction : théorie et pratique*. Actes du colloque international intitulé Traduction humaine, traduction automatique, interprétation, Tunis les 28,29 et 30 septembre 2000.
- Vandaele (S.), 2002 : « Métaphores conceptuelles en traduction médicale et cohérence », dans *TTR*, volume XV, n° 1, p. 223-239.

Pierre J.L. Arnaud,
CRTT, Université Lumière,
Lyon,
France,
pierre.arnaud@univ-lyon2.fr.

Détecter, classer et traduire les métonymies (anglais et français)

Résumé

L'étude porte sur les métonymies présentes dans un corpus de presse française et anglaise. On examine d'abord les critères d'inclusion, de nombreuses occurrences étant marginales. Une typologie récente des métonymies, celle de Peirsman et Geeraerts (2006), est adaptée, avec 54 catégories résultantes, et un échantillon de 100 occurrences dans chaque langue est confronté à ces catégories. Il apparaît des cas de correspondance évidente, mais aussi des catégorisations moins nettes ou ambiguës justifiant une approche prototypiste, ainsi qu'un résidu d'inclassables. Un effectif de 54 catégories est donc insuffisant. Enfin, les divers cas de correspondance anglais > français sont répertoriés et discutés.

Mots-clés : métonymie ; classification ; corpus ; traduction.

Introduction

Depuis une vingtaine d'années, la métonymie a fait l'objet d'une activité de recherche intense, principalement dans le cadre de la linguistique cognitive. La réflexion a ainsi été considérablement approfondie, même s'il faut bien dire que, fondamentalement, on ne sait toujours pas de façon assurée sur quels mécanismes repose ce phénomène et ce qui le motive. Il existe cependant un consensus parmi les cognitivistes : ainsi, la métonymie est, au moins depuis Lakoff et Johnson (1980), considérée non pas simplement comme un phénomène lexical (et *a fortiori* rhétorique), mais comme un phénomène conceptuel, impliqué dans la cognition. Sous l'influence des idées de Langacker (1987) sur les concepts et leurs domaines, la notion de contiguïté, déjà évoquée par Aristote à propos de la mémoire, a chez de nombreux auteurs cédé le pas à la notion de relation intradomaine, malgré les faiblesses de la notion de domaine (sur ce dernier point, voir par exemple Riemer 2003).

Beaucoup de publications ont une orientation théorique, utilisant des exemples fabriqués ou repris d'un article à l'autre, le plus célèbre étant sans doute celui du sempiternel sandwich au jambon de Nunberg (1978). D'autres recherches ont pris en compte des exemples relevés en corpus, mais dans ce cas, la réflexion a porté sur un type particulier de métonymie, par exemple celles où la source est un nom de pays (Lecolle 2001, 2002) ou bien un nom d'organe (Hilpert 2006 sur *eye*), parfois

avec le but d'entraîner un algorithme à la détection des métonymies (Markert et Nissim 2002).

La recherche présentée ici a eu pour principe de partir des données collectées. Pour ce faire, deux corpus de presse, l'un français et l'autre anglais, ont été examinés pour en extraire toutes les occurrences apparaissant comme métonymiques. Cette extraction fait l'objet de la section suivante. Nous examinerons ensuite la répartition des types issus de cette étape par rapport à une typologie récemment publiée. Enfin, la traduction des types du corpus anglais sera considérée.

1 Détecter les métonymies

1.1 Les corpus

Partant du principe que les journalistes, qui écrivent sous contrainte de temps et recherchent souvent un style pittoresque et imagé, font un abondant usage de phraséologies et de tropes ritualisés, un corpus de presse a été assemblé pour chaque langue afin de collecter des métonymies. Les deux corpus se composent de quotidiens, gratuits ou payants, et de magazines d'actualité et féminins. Étant donné l'impossibilité actuelle d'une détection automatique, les textes furent examinés visuellement, les occurrences étant reportées dans une base de données. Dans ce stade initial, les occurrences de métaphores ont également été relevées. S'efforcer de prendre en compte toutes les occurrences impose, tautologiquement, la détection de *toutes* les occurrences malgré les obstacles, mais aussi le rejet d'occurrences qui ne constituent pas véritablement des cas de métonymie.

1.2 Détection positive : toutes les métonymies

L'examen des textes étant visuel, une attention soutenue est nécessaire, mais ceci n'est pas propre à la métonymie. Parmi les obstacles spécifiques, le plus redoutable est constitué par certaines occurrences lexicalisées ou catachrèses, qui ont perdu toute visibilité. Il y a ainsi un risque de sauter des occurrences telles que celles d'*argent* (métal > monnaie), résultats de changements si incrustés (*entrenched*) dans la langue qu'elles passent inaperçues. La dimension diachronique est donc présente, que ce soit dans la polysémie comme pour l'exemple ci-dessus, ou dans le changement total de sens, comme pour *reek* (« smoke » > « stench »). Les occurrences vives (originales) sont évidemment plus facilement repérables par la rupture d'isotopie qu'elles entraînent, mais les métonymies ne se laissent pas si facilement répartir en catachrèses d'un côté et métonymies vives de l'autre. Il existe, tout particulièrement dans le discours de presse, des états intermédiaires où une occurrence semble correspondre à un type déjà rencontré, sans que ce type soit répertorié par les dictionnaires généraux, par exemple *projet* « ensemble de bâtiments », sens absent du *Petit Robert* (2009), rattachable à la catégorie CAUSE > RÉSULTAT (et sans doute un anglicisme).

La prise en compte de toutes les occurrences impose en outre de ne pas se limiter aux seuls mots, simples ou composés comme *hothead* (« excité », synecdoque), mais aussi d'intégrer les phraséologies. En effet, si les métonymies sont rarement filées sur plusieurs propositions comme les métaphores peuvent l'être, elles peuvent néanmoins occuper plusieurs mots graphiques.

1.3 Détection négative : rien que les métonymies

La métonymie est un phénomène bien plus complexe que la métaphore et, devant une occurrence apparemment métonymique, un examen serré est parfois nécessaire afin de confirmer ce statut. Un premier cas résulte de la succession de plusieurs métonymies, comme dans *denim* « 1) tissu > 2) vêtements > 3) style de vie caractérisé par le port de tels vêtements », ou *ardoise* « 1) roche > 2) support d'écriture > 3) somme à régler dont le montant est écrit sur une ardoise » ; il est donc nécessaire de faire le parcours en sens inverse pour s'assurer de sa nature. Plus gênantes sont les « métaphonymies », ainsi dénommées par Goossens (1995) dans un article qui relança l'intérêt pour ce phénomène. On savait en effet depuis longtemps que la métaphore et la métonymie pouvaient se combiner, mais l'attention portée à ce fait par la linguistique cognitive (Barcelona 2000, Dirven et Pörings 2003) a permis d'approfondir la question. Un exemple de superposition est *ardoise*, « dette », car la situation qui a donné naissance à la dette présente une analogie avec celle où la note, de restaurant ou autre, est inscrite sur une ardoise. Le problème est que de nombreux cas sont moins nets et qu'on peut finir par se persuader que toute métaphore comporte une part de métonymie. Chaque cas de mixité potentielle réclame donc une réflexion approfondie. Prenons *bémol* « chose ou événement de nature à réduire l'aspect positif d'une autre chose ou d'un autre événement ». On perçoit bien sûr immédiatement la métaphore par rapport au signe du solfège sur le « sème » [ABAISSER]. Mais le bémol du solfège étant un signe, avons-nous en plus une métonymie symbole > symbolisé ? Autre exemple : « *We can't go there with the focus of trying to protect the lead* ». *Focus* est une métaphore du domaine de l'optique, mais également une métonymie, puisque la cible n'est pas le foyer, mais l'objet visé. L'attitude adoptée dans la présente recherche est de compter comme unités mixtes de telles occurrences, et de prendre en compte leur côté métonymique pour le classement. Une anecdote illustrera les difficultés d'analyse : j'ai soumis l'exemple italien de Koch (2004), *Che casino!* « Quel bordel ! » à quatre linguistes, et deux y ont vu une métonymie et les deux autres une métaphore...

Par ailleurs, les vues de la linguistique cognitive, en ayant fait de la métonymie un phénomène conceptuel, nous ont éloignés d'une définition comme trope. C'est ainsi qu'on en est arrivé à considérer comme métonymiques de nombreux faits d'expression indirecte ou de représentation. On peut ainsi hésiter devant un énoncé comme :

[1] *Can you lend me your car?*

Il s'agit d'un acte de langage, dans lequel on ne pose pas réellement une question sur la capacité mais on demande une action, donc d'interprétation non littérale (Panther et Thornburg 1998).

Dans la même veine, on a pu voir de la synecdoque, donc de la métonymie, dans l'appartenance catégorielle, matérialisée en langue par la relation hypo-/hyperonyme et exploitée dans l'anaphore hyperonymique, comme dans :

[2] *The V-shaped rock structure was built [...] by Britons who used the tidal flow of the estuary to trap THE FISH. Once herded into the contraption, THE CREATURES could be simply scooped out with nets.* (Metro, 17/03/09)

La synecdoque peut aussi être vue comme participant à la restriction ou à l'élargissement du sens, comme dans respectivement *groupe* « ensemble > ensemble de compagnies appartenant au même propriétaire » et *gentleman* « *man of noble birth* > man ». Nous n'irons pas jusque-là.

Un doute de nature différente est introduit par Croft (1993), qui prend l'exemple *this book is heavy/this book is a history of Iraq*, dont il dit qu'il n'est pas clair qu'il y ait référence à deux entités distinctes. Comme les éléments « profilés » dans chaque domaine sont hautement intrinsèques, il s'agit plutôt de deux « facettes » d'un même concept (cf. Croft et Cruse 2004 : 116sq.). Devant l'impossibilité de décider objectivement de ce qui constitue une facette, ce genre d'occurrence ne sera pas rejeté. Par contre, bien que la conversion N > V implique une compression de l'information, comme dans [*The pub*] *screens major sporting occasions*, elle n'a été prise en compte que dans les cas où elle repose sur une véritable métonymie comme dans ce dernier exemple.

Enfin, on pourrait aller jusqu'à voir de la métonymie dans (x) *the loss of my wallet put me in a bad mood* > (y) *the fact of having no money left* (Panther et Thornburg 2004). En fait, Panther et Thornburg tracent ici une limite, car selon eux le lien entre x et y est nécessaire, alors que le lien cible-source dans une métonymie est toujours contingent. Le problème est que le discours de presse regorge d'« inférences invitées », sans doute pour augmenter l'intérêt du lecteur, et qu'il n'est pas toujours si simple de décider si un lien est nécessaire ou contingent. L'exemple suivant semble cependant clair : (dans un contexte de conseils pour réduire ses dépenses automobiles) « *Press lightly on the accelerator.* » Non seulement le lien « appui sur accélérateur > consommation d'essence » est nécessaire, mais encore l'énoncé a du sens littéral. Une telle occurrence ne sera donc pas prise en compte.

N'ont pas été incluses non plus les périphrases journalistiques comme *the Emerald Isle*, *la planète rouge*, *les billets verts*, ou anaphoriques, comme *the black stuff* « Guinness », malgré leur référence indirecte, car le lien cible-périphrase est un lien de définition description et non de substitution par associé.

Enfin, si la métonymie, en reposant sur un lien implicite entre cible et source, comprime l'information, c'est aussi le cas de l'ellipse ou de la détermination par la situation, qui sont des phénomènes d'une autre nature. Il ne semble donc pas que des occurrences comme « C'est qui, la chèvre ? » (= « C'est pour qui, la pizza au fromage de chèvre ? » — Lecolle 2001) doivent être retenues.

Une définition restrictive de la métonymie a donc été appliquée dans le but d'éviter la prolifération des exemples, mais on aura compris que, dans beaucoup de cas, des décisions d'inclusion doivent être prises en l'absence de critères fermes. Il est alors essentiel de garder des notes pour assurer la cohérence du classement.

1.4 Données et statistiques

La même procédure a été suivie pour chaque langue. Dans un premier temps, 1 000 occurrences tropiques ont été intégrées à la base de données. Les occurrences ont été classées en métonymiques, métaphoriques, métaphtonymiques, et indécidables. Le tableau 1 montre la répartition de ces catégories, et notamment, la domination en discours de la métaphore sur la métonymie.

| | français | | anglais | |
|----------------|----------|------|----------|------|
| | effectif | | effectif | |
| métonymies | 216 | 21,6 | 202 | 20,2 |
| métaphores | 717 | 71,7 | 742 | 74,2 |
| métaphtonymies | 55 | 5,5 | 54 | 5,4 |
| inclassables | 12 | 1,2 | 2 | 0,2 |
| total | 1 000 | 100 | 1 000 | 100 |

Tableau 1 : Effectifs sur 1 000 occurrences

Ces statistiques sur les occurrences de tous types une fois établies, la collecte a été poursuivie sur les seules occurrences métonymiques et métaphtonymiques (sans distinction), jusqu'à concurrence de 500, puis le regroupement en types a été effectué. Pour les métaphtonymies, c'est évidemment la seule composante métonymique qui a compté à ce stade. Le nombre de types correspondants s'établit à 353 pour le français et 285 pour l'anglais. La différence, qui est significative ($\chi^2 = 4,22$ [Yates], $p = 0,04$), s'explique sans doute par la structure très répétitive des articles de journaux anglais comparés aux articles français. Il est utile de rappeler ici que le but de cette recherche est principalement d'étudier la catégorisation des métonymies et que catégoriser les types suffit à catégoriser les occurrences. Le tableau 2 (page suivante) montre les types les plus fréquents dans les deux corpus.

La littérature ne manque pas de listes de typologies de la métonymie (voir Kövecses et Radden 1998; Blank 1999; Arnaud 2003 : 87-92, 2006), mais beaucoup sont partielles, limitées aux catégories les plus fréquentes citées comme exemples. Parmi les typologies visant à une plus grande exhaustivité, on peut citer celle du dictionnaire de Morier (1961/1989 : 768sq.) avec 28 catégories et, dans le cadre théorique de la linguistique cognitive, celle de Kövecses et Radden (1998), subdivisée de façon complexe et comportant une trentaine de catégories. Un emploi de cette liste ayant produit précédemment un résidu d'inclassables (Arnaud 2006), on a préféré utiliser une typologie originaire de la même mouvance théorique, mais plus récente, celle de Peirsman et Geeraerts (2006), qui ont compilé leur liste à partir de publications classiques.

| français | | | anglais | | |
|----------|--------------------------|-----------|---------|---|-----------|
| rang | métonymie | fréquence | rang | métonymie | fréquence |
| 1 | gauche (= politique) | 12 | 1 | <i>sex</i> (= <i>sexual activity</i>) | 17 |
| 2 | argent (= monnaie) | 7 | 2 | <i>market(s)</i> (<i>finance</i>) | 11 |
| | cadres (= chefs) | | 3 | <i>cash</i> | 10 |
| 4 | marque (= entreprise) | 6 | 4 | <i>bank</i> | 9 |
| 5 | album (= œuvre) | 5 | 5 | <i>album</i> (= <i>work</i>) | 7 |
| | droite (politique) | | | <i>side</i> (= <i>team</i>) | |
| | grève | | 7 | <i>board</i> (= <i>administrators</i>) | 6 |
| 8 | banque | 4 | | <i>charity</i> (= <i>ngo</i>) | |
| | chaîne (= entreprise TV) | | | <i>high street</i> | |
| | crèche | | | <i>bill</i> (= <i>sum</i>) | |

Tableau 2 : Types les plus fréquents (sur 500 occurrences)

Alors que certaines catégories étaient régulièrement citées par les auteurs qu'ils ont consultés, d'autres n'apparaissaient que dans une seule source ou référaient à des entités très spécifiques. Pour la présente recherche, un certain nombre de modifications ont été apportées à la liste. Certaines catégories ont été doublées d'une catégorie « b » correspondant au mouvement inverse lorsque ceci n'était pas le cas chez Peirsman et Geeraerts. D'autres catégories, comme ENTITÉ > ENTITÉ ADJACENTE, n'avaient logiquement pas besoin d'une orientation inverse. Enfin, la catégorie PARTICIPANT > AUTRE PARTICIPANT a été introduite; chez Peirsman et Geeraerts, elle ne figure que comme supercatégorie couvrant des relations plus spécifiques comme PRODUCTEUR > PRODUIT, mais on a voulu anticiper sur l'apparition de types non couverts par les catégories spécifiques. Il en a été de même pour ENTITÉ > ENTITÉ ADJACENTE. Pour les raisons exposées plus haut, les catégories POTENTIEL ET ACTUEL (*Can you see him?*) et HYPONYME ET HYPERONYME (*la pilule*) n'ont pas été retenues.

Les intitulés ont été traduits pour la liste ci-dessous, et chaque fois que les auteurs fournissaient un ou plusieurs exemples, un exemple a été traduit dans la mesure du possible. Le terme *procès* correspond à *action/event/process* chez eux. Telle qu'elle est, la typologie comporte des catégories situées à des niveaux divers de généralité. Les auteurs mentionnent ce fait et surtout insistent sur le fait que leur typologie n'a pas de prétention à l'exhaustivité. Leur but en la compilant n'était d'ailleurs aucunement de classer des occurrences de discours, mais d'explorer la possibilité d'une catégorisation « de haut niveau », abstraite, objectif différent de celui de la présente recherche. La typologie s'établit comme suit :

- | | | |
|-----|------------------------|---|
| 1. | PARTIE SPATIALE > TOUT | un Peau-Rouge |
| 1b. | TOUT SPATIAL > PARTIE | l'Amérique (= les USA) |
| 2. | CONTENEUR > CONTENU | (boire) un verre |
| 2b. | CONTENU > CONTENEUR | le lait s'est renversé (= la bouteille) |

3. LOCALISATION > LOCALISÉ tout le théâtre a applaudi
3b. LOCALISÉ > LOCALISATION université (communauté > bâtiment)
4. ENTITÉ > ENTITÉ ADJACENTE une table ronde (= les gens assis autour
de la table)
5. VÊTEMENT > PERSONNE une vieille perruque (= une personne
5b. PERSONNE > VÊTEMENT âgée)
6. VÊTEMENT > PARTIE DU CORPS *Sohle* (semelle > plante du pied)
6b. PARTIE DU CORPS > VÊTEMENT
7. OBJET > MATIÈRE de l'hermine
7b. MATIÈRE > OBJET un carton
8. PARTIE TEMPORELLE > TOUT *Morgen* (matin > demain)
8b. TOUT TEMPOREL > PARTIE
9. TEMPS > ENTITÉ 14-18 a changé la face du monde
9b. ENTITÉ > TEMPS
10. ANTÉCÉDENT > CONSÉQUENT Grec *phobos* (fuite > peur)
10b. CONSÉQUENT > ANTÉCÉDENT
11. SOUS-ÉVÉNEMENT Corriger des copies (= corriger et
> événement complexe noter des copies)
11b. ÉVÉNEMENT COMPLEXE > SOUS-ÉVÉNEMENT
12. PROCÈS > AGENT L'assistance (= le public)
12b. AGENT > PROCÈS *To author a book*
13. PROCÈS > PATIENT Des achats
13b. PATIENT > PROCÈS
14. PROCÈS > LOCALISATION Un passage (ex. : entre obstacles)
14b. LOCALISATION > PROCÈS L'Amérique ne voulait pas un deuxième
Pearl Harbor.
15. PROCÈS > TEMPS À la fenaison
15b. TEMPS > PROCÈS *To summer in Paris*
16. PROCÈS > INSTRUMENT *Andenken* (= un souvenir)
16b. INSTRUMENT > PROCÈS *To shampoo one's hair*

- | | |
|---|--|
| 17. État > LOCALISATION | <i>Gehorsam</i> (obéissance > prison) |
| 17b. LOCALISATION > ÉTAT | |
| 18. CAUSE > EFFET | Ouvrir les prisons (libérer les prisonniers); une lâcheté |
| 18b. EFFET > CAUSE | <i>To empty a glass</i> , une lumière (une lampe) |
| 19. PARTICIPANT > AUTRE PARTICIPANT (autre que 17-21) | |
| 20. CONTRÔLEUR > CONTRÔLÉ | Ozawa a donné un mauvais concert (l'orchestre dirigé par O.) |
| 20b. CONTRÔLÉ > CONTRÔLEUR | La Mercedes est arrivée. |
| 21. POSSÉDÉ > POSSESSEUR | La courte paille (celui qui l'a tirée) |
| 21b. POSSESSEUR > POSSÉDÉ | C'est Paul (en montrant une boisson) |
| 22. PRODUCTEUR > PRODUIT | Une Ford, un Picasso |
| 22b. PRODUIT > PRODUCTEUR | Un coucou (oiseau) |
| 23. LOCALISATION > PRODUIT | Du Bordeaux |
| 23b. PRODUIT > LOCALISATION | Madeira (bois > île) |
| 24. INSTRUMENT > RÉSULTAT | <i>A whistle</i> (sifflet > coup de sifflet) |
| 24b. RÉSULTAT > INSTRUMENT | |
| 25. CARACTÉRISTIQUE > ENTITÉ | C'est une beauté. |
| 25b. ENTITÉ > CARACTÉRISTIQUE | |
| 26. COLLECTION > INDIVIDU | <i>Ein Frauenzimmer</i> (le sexe féminin > une bonne femme) |
| 26b. INDIVIDU > COLLECTION | |
| 27. OBJET > QUANTITÉ | |
| 27b. QUANTITÉ > OBJET | Un quart (des sardines en boîte) |
| 28. FACTEUR CENTRAL > INSTITUTION | La presse |
| 28b. INSTITUTION > FACTEUR CENTRAL | |

Le nombre de catégories est donc de 28, ou 54 si on prend en compte les inversions de direction.

Pour maintenir la tâche dans des limites raisonnables, un échantillon aléatoire de 100 types de chaque langue a ensuite été extrait des deux listes et confronté à la typologie. L'analyse a été effectuée à plusieurs reprises, jusqu'à ce qu'il subsiste le

moins possible d'hésitations, mais comme on va le voir, il en reste. Plusieurs cas se sont présentés. Dans le plus simple, l'adéquation d'une catégorie au type examiné est apparue immédiatement et sans hésitation, comme pour :

- [3] une fondation 13. PROCÈS > PATIENT
- [4] la gendarmerie (= des gendarmes) 26. COLLECTION > INDIVIDU
- [5] les premiers pas sur Mars (= le fait d'atteindre) 11. SOUS-événement > événement complexe
- [6] *the catwalks* (= *the world of fashion*) 28. FACTEUR CENTRAL > INSTITUTION
- [7] *metal* (= *a car*) 7b. MATIÈRE > OBJET
- [8] *a pint* (= *a drink*) 27b. QUANTITÉ > OBJET

Dans d'autres cas, la catégorisation a nécessité plusieurs réexamens, sans que le résultat soit entièrement convaincant :

- [9] un comité d'établissement (= une séance du comité) 12b. AGENT > PROCÈS ?
- [10] le tableau (= la situation) 2. CONTENEUR > CONTENU ?
- [11] *admissions* (= *number of admitted people*) 27. OBJET > QUANTITÉ ?
- [12] *a big shot* (*an important person*) 12. PROCÈS > AGENT ?

Je ne suis pas le premier à éprouver des difficultés dans une telle tâche, et Riemer (2003), par exemple, signale l'indétermination de nombreuses métonymies. Deux explications peuvent être avancées dans le cas présent. La plus simple est que je n'aurais pas bien compris la typologie de Peirsman et Geeraerts (2006) ou pas su l'appliquer, et ceci vaut pour les cas suivants aussi. Une explication plus positive repose sur l'approche prototypiste qui est celle de ces auteurs eux-mêmes, et que je partage. Dans cette perspective, il y a des métonymies qui seront au cœur de la catégorie concernée, et d'autres, situées à la périphérie, seront parfois problématiques.

Dans la même perspective, les doubles catégorisations résultent de métonymies équidistantes de deux cœurs de catégories :

- | | |
|--|---------------------------------------|
| [13] <i>les paillettes</i> (= une vie de plaisirs) | 1. PARTIE SPATIALE > TOUT ? |
| | 16b. INSTRUMENT > PROCÈS ? |
| [14] <i>a jet</i> (an aircraft) | 19. PARTICIPANT > AUTRE PARTICIPANT ? |
| | 25. CARACTÉRISTIQUE > ENTITÉ ? |

À l'extrême, on a les métonymies pour lesquelles aucune solution n'apparaît :

- [15] l'édition (= secteur d'activité)
- [16] une nounou
- [17] la technologie (= une technique spécifique)

[18] *the circulation* (of a newspaper, = the number of copies sold)

[19] *blue grass* (music)

[20] *swear* (use offensive language)

À ce propos, il convient de se rappeler l'avertissement de Peirsman et Geeraerts sur le caractère non exhaustif de leur typologie. Que certaines catégories manquent apparaît parfois quand on examine l'ensemble des données au-delà des deux échantillons de 100, ainsi :

Catalogue > objets; liste > personnes; portefeuille > valeurs mobilières; palette > ensemble de couleurs; agenda > choses à faire; calendrier > déroulement d'événements; programme > suite d'émissions; répertoire > œuvres; argumentaire > arguments; cahier des charges > spécifications; curriculum > *someone's past*; *order book* > *orders*.

De telles métonymies sont à la marge de la catégorie CONTENEUR > CONTENU; on note en effet que les conteneurs n'y sont pas de bons représentants de leur catégorie et que certains des contenus ne sont pas concrets. Il apparaît donc souhaitable de sous-catégoriser CONTENEUR > CONTENU.

Une autre catégorie apparaît dans les métonymies suivantes : *argent* > 2^e place dans une compétition sportive; *enseigne* > entreprise; *la quille* > le départ en retraite; *lambda* > moyen, quelconque; *horloge qui tourne* > temps qui passe. Elle n'est pas absente de la littérature et a été repérée entre autres par Ullmann (1969 : 285-286) qui l'a étiquetée « le symbole pour la chose symbolisée ». Enfin, comme nous l'avons vu, la place de ENTITÉ > ENTITÉ ADJACENTE dans la typologie de Peirsman et Geeraerts (2006) n'est pas claire, et PARTICIPANT > AUTRE PARTICIPANT a été rajouté. Il s'avère à l'usage que PARTICIPANT > AUTRE PARTICIPANT est un intitulé trop vague qui rend la catégorie excessivement accueillante.

Le tableau 3 montre le résultat du classement des deux échantillons de 100 types métonymiques. Il est clair que la typologie de Peirsman et Geeraerts, même modifiée pour la présente recherche, n'échappe pas à ce que Blank (1999) écrivait à propos de travaux antérieurs : « *typologies of metonymy [...] lack consistency in that there are usually metonymies that do not fit into any category.* » Il s'agit non seulement d'un problème de granularité (CONTENEUR > CONTENU à sous-catégoriser), mais aussi de catégories manquantes. On peut conclure de cette section qu'on ne peut se contenter de 54 catégories puisqu'un tel effectif est déjà mis en échec par des échantillons de métonymie ne dépassant pas 100 unités.

| | français | anglais |
|-------------------------|----------|---------|
| catégorisation claire | 61 | 62 |
| catégorisation douteuse | 13 | 20 |
| double catégorisation | 14 | 5 |

| | | |
|---|-----|-----|
| inclassables | 7 | 11 |
| appartenance claire à catégorie manquante | 5 | 2 |
| total | 100 | 100 |

Tableau 3 : Résultats du classement des métonymies

2 Traductions

Comment les métonymies se traduisent-elles ? Si la métonymie est un phénomène conceptuel avant d'être linguistique, on peut faire l'hypothèse que les locuteurs de langues différentes recourent aux mêmes catégories de métonymies conceptuelles, mais on ne peut exclure que les langues privilégient ou au contraire soient réfractaires à certaines catégories.

Plusieurs cas sont théoriquement possibles : si une métonymie est lexicalisée, elle a généralement un équivalent lexicalisé dans la langue cible, mais celui-ci peut comprendre une métonymie d'une catégorie différente, ou bien n'avoir rien de métonymique. Le traducteur peut également être contraint de trouver une solution originale en l'absence d'équivalent préexistant. Si la métonymie est originale, l'effet de style prend de l'importance et une traduction littérale peut préserver la métonymie et donc l'effet. Il peut également arriver qu'une telle traduction soit impossible et que l'effet soit perdu ou doive être compensé par d'autres moyens.

Pour voir ce qu'il en est de ces considérations *a priori*, les 100 métonymies anglaises de l'échantillon ont été traduites. Ont été pris en compte le caractère lexicalisé ou original de la métonymie et la nature de l'équivalent français. Pour ce qui est de l'opposition lexicalité ~ originalité, nous avons vu que ceci n'est pas une question de « tout ou-rien » ; on a cependant eu recours à un classement binaire pour la présente étape : lorsque le *Concise Oxford English Dictionary* (2004) indiquait une sous-entrée ou un changement de sens métonymiques, on a considéré l'exemple comme lexicalisé. La domination dans le discours de presse de la métonymie lexicalisée est flagrante : 91 des 100 unités de l'échantillon s'avèrent être dans ce cas.

Les exemples suivants montrent les différentes combinaisons :

- (a) métonymie lexicalisée > métonymie française lexicalisée, traduction directe et de la même catégorie

[21] *alcohol* > alcool (boissons alcoolisées)

[22] *the ring* > le ring (la boxe)

Noter que l'équivalent français peut être une variante :

[23] *be history* > être de l'histoire ancienne

- (b) métonymie lexicalisée > métonymie française lexicalisée différente, mais de la même catégorie
 - [24] *a duty* > un droit (une taxe; noter l'opposition de modalité)
 - [25] *gladiators* > des s̄partiates
- (c) métonymie lexicalisée > métonymie française lexicalisée d'une catégorie différente
 - [26] *a date* > un rendez-vous (= une personne)
 - [27] *a dealership* > une concession
- (d) métonymie lexicalisée, équivalent français non métonymique
 - [28] *the boardroom* > le conseil d'administration, les dirigeants
 - [29] *an effort* > un tir (football)
- (e) métonymie originale, traduisible par une métonymie
 - [30] *anoraks* > des anoraks (des personnes)
 - [31] *catwalks, runways* > les podiums (le monde de la mode)
- (f) métonymie originale, traduction non métonymique
 - [32] *platforms* > chaussures à semelles compensées
 - [33] *the armalite* > le fusil-mitrailleur

Sur les 91 métonymies lexicalisées de l'échantillon, 35 sont traduisibles par une métonymie française directement équivalente et 15 par une métonymie différente. La métonymie ne constitue donc quantitativement pas un obstacle considérable au travail du traducteur.

Conclusion

Si la traduction n'est apparemment pas un problème spécifique posé par la métonymie, il n'en manque pas d'autres. La linguistique cognitive a sans doute contribué à compliquer la situation en considérant la métonymie comme un phénomène conceptuel, ce qui la rend plus envahissante, et l'absence de consensus sur les limites de la métonymie impose, on l'a vu, des décisions d'inclusion parfois délicates.

Il n'existe pas actuellement de typologie de la métonymie qui soit directement utilisable, et cela même dans la perspective prototypiste qui s'impose au vu des données. La question de la granularité est au centre des problèmes de catégorisation, et nous avons vu qu'un effectif de 54 catégories ne donne pas un résultat satisfaisant. Le constat qu'Henry (1971 : 18) faisait quant à l'incapacité de la rhétorique ancienne « de dresser un relevé exhaustif des prétendues espèces de métonymie et de synecdoque » reste valable malgré les progrès théoriques de ces 25 dernières années. Certes, classer n'est pas expliquer, mais classer aide à expliquer.

Bibliographie

- Arnaud (P.), 2003 : *Les Composés timbre-poste*, Lyon, Presses Universitaires de Lyon.
- Arnaud (P.), 2006 : « Phénomènes de sous-spécification sémantique, représentation des unités lexicales et prise en compte du contexte », dans Blampain (D.), Thoiron (P.) et Van Campenhoudt (M.), éd., *Mots, termes et contextes*, Paris, Éditions des Archives Contemporaines, p. 33-43.
- Barcelona (A.), dir., 2000 : *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*, Berlin, Mouton-de Gruyter.
- Blank (A.), 1999 : « Co-presence and succession: A cognitive typology of metonymy », dans Panther (K.-U.) et Radden (G.), dir., *Metonymy in Language and Thought*, Amsterdam, Benjamins, p. 169-191.
- COD = *Concise Oxford English Dictionary*, 11^e édit., 2004, Soanes (C.) and Stevenson (A.), dir., Oxford, Oxford University Press.
- Croft (W.), 1993 : « The role of domains in the interpretation of metaphors and metonymies », dans *Cognitive Linguistics*, vol. 4, p. 335-370.
- Croft (W.) et Cruse (D.A.), 2004 : *Cognitive Linguistics*, Cambridge, C.U.P.
- Dirven (R.) et Pörings (R.), dirs., 2003 : *Metaphor and Metonymy in Comparison and Contrast*, Berlin, Mouton-de Gruyter.
- Goossens (L.), 2003 : « Metaphonymy: The interaction of metaphor and metonymy in expressions for linguistic action », dans Dirven (R.) et Pörings (R.), dir. *Metaphor and Metonymy in Comparison and Contrast*, Berlin, Mouton-de Gruyter, p. 349-377.
- Henry (A.), 1971 : *Métonymie et métaphore*, Paris, Klincksieck.
- Hilpert (M.), 2006 : « Keeping an eye on the data: Metonymies and their patterns », dans Stefanowitsch (A.) et Gries (S.T.), dir., *Corpus-based Approches to Metaphor and Metonymy*, Berlin, Mouton-de Gruyter, p. 123-52.
- Koch (P.), 2004 : « Metonymy between pragmatics, reference, and diachrony », dans *Metaphorik.de*, www.metaphorik.de/, mise en ligne décembre 2004.
- Kövecses (Z.) et Radden (G.), 1998 : « Metonymy: Developing a cognitive linguistic view », dans *Cognitive Linguistics*, vol. 9, p. 37-77.
- Lakoff (G.) et Johnson (M.), 1980 : *Metaphors we Live by*, Chicago, University of Chicago Press.
- Langacker (R.W.), 1987 : *Foundations of Cognitive Grammar. Vol. 1: Theoretical prerequisites*, Stanford, Stanford University Press.
- Lecolle (M.), 2001 : « Métonymie dans la presse écrite: entre discours et langue », dans *Travaux Neuchâtelois de Linguistique*, vol. 34/35, p. 153-170.
- Lecolle (M.), 2002 : « Personnifications et métonymies dans la presse écrite: comment les différencier? », dans *Semen, 15, Figures du discours et ambiguïté*, Collection Annales Littéraires de l'Université de Franche-Comté, Besançon, Presses Universitaires de Franche-Comté, p. 97-112.
- Markert (K.) et Nissim (M.), 2002 : « Metonymy resolution as a classification task », dans *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, Philadelphia, Penn., 6-7 July 2002*, p. 204-213.
- Morier (H.), 1961 : *Dictionnaire de poésie et de rhétorique*, Paris, PUF [édition citée: 1989].

- Nunberg (G.D.), 1978 : *The Pragmatics of Reference*, Bloomington, Indiana University Linguistics Club.
- Panther (K.-U.) et Thornburg (L.L.), 1998 : « A cognitive approach to inferencing in conversation », dans *Journal of Pragmatics*, vol. 30, p. 755-769.
- Panther (K.-U.) et Thornburg (L.L.), 2004 : « The role of conceptual metonymy in meaning construction », dans *Metaphorik.de*, www.metaphorik.de/, mise en ligne juin 2004.
- Peirsman (Y.) et Geeraerts (D.), 2006 : « Metonymy as a prototypical category », dans *Cognitive Linguistics*, vol. 17, p. 269-316.
- Petit Robert 2009 = *Le Nouveau Petit Robert: Dictionnaire alphabétique et analogique de la langue française, 2009, Version électronique*, Rey-Debove (J.) et Rey (A.), dir.
- Riemer (N.), 2003 : « When is a Metonymy no longer a Metonymy? », dans Dirven (R.) et Pörings (R.), dir., *Metaphor and Metonymy in Comparison and Contrast*, Berlin, Mouton-de Gruyter, p. 379-406.
- Ullmann (S.), 1969 : *Précis de sémantique française*, 4^e édit., Berne, Francke.

Pedro Mogorrón Huerta,
Universidad de Alicante,
Alicante,
Espagne,
pedro.mogorron@ua.es.

Compétence phraséologique et traitement des expressions figées dans les dictionnaires

Résumé

À partir d'une sélection d'une vingtaine d'expressions figées (EF) en français et en espagnol qui appartiennent à différents niveaux de langue, nous désirons analyser la compétence phraséologique des usagers et le traitement donné par les dictionnaires monolingues, phraséologiques et bilingues à ces EF.

Mots-clés : expressions figées ; phraséologie ; dictionnaire bilingue ; équivalence.

Introduction : délimitation du travail

De nos jours, dans de nombreux pays, des unités de recherche analysent les expressions figées¹. Cependant, le recensement, le traitement, la classification de ces séquences, ainsi que l'analyse contrastive et la traduction exhaustive de ces EF entre deux ou plusieurs langues restent encore à faire dans de nombreux cas². À partir d'une sélection d'une vingtaine d'expressions figées en français et en espagnol d'usage fréquent³ qui appartiennent à différents niveaux de langue, et afin de nous permettre de vérifier si la compétence phraséologique est la même dans chaque niveau de langue, nous analyserons :

- la présence et le traitement de ces EF dans les principaux dictionnaires monolingues et phraséologiques ;
- la compétence phraséologique des étudiants ;
- la traduction des expressions figées ;
- le traitement donné par les dictionnaires bilingues à ces expressions ;

¹ Que nous noterons *EF* dans cet article.

² Ces observations apparaissent dans le cadre des expériences et des réflexions menées lors de l'élaboration d'une base de données de séquences figées multilingue à l'Université d'Alicante.

³ Pour vérifier la fréquence d'emploi de ces expressions, nous avons cherché sur Internet le nombre de fois où elles apparaissent dans google.fr/google.es. De même, il faudra réaliser un travail équivalent avec des EF non fréquentes pour observer leur traitement phraséologique dans les mêmes dictionnaires utilisés ici.

- la recherche de la meilleure équivalence. (Nous analyserons la problématique des nombreuses expressions figées qui ont plusieurs parasynonymes comme *avoir peur, être distrait*).

1 Analyse des EF

Nous allons analyser dans le cadre de cet article une vingtaine d'EF en français et en espagnol que nous reproduisons ci-dessous dans leur intégralité :

EF françaises : 1f) adorer le veau d'or ; 2f) franchir le Rubicon ; 3f) s'endormir sur ses lauriers ; 4f) tuer la poule aux œufs d'or ; 5f) se croire sorti de la cuisse de Jupiter ; 6f) prendre la clé des champs ; 7f) être dans les nuages ; 8f) arrondir les angles ; 9f) mettre la charrue avant les bœufs ; 10f) tirer les marrons du feu ; 11f) avoir une mémoire d'éléphant ; 12f) se serrer la ceinture ; 13f) n'en avoir rien à foutre de ; 14f) baisser son froc ; 15f) faire de l'œil ; 16f) fermer sa boîte ; 17f) fermer son bec ; 18f) faire dans son froc.-

EF espagnoles : 1e) *adorar el becerro de oro* ; 2e) *pasar el Rubicón* ; 3e) *dormirse en los laureles* ; 4e) *matar la gallina de los huevos de oro* ; 5e) *creerse descendiente de la pata del Cid* ; 6e) *tomar las de Villadiego* ; 7e) *estar en las nubes* ; 8e) *limar (las) asperezas* ; 9e) *empezar la casa por el tejado* ; 10e) *sacar(le) las castañas del fuego a alguien* ; 11e) *tener memoria de elefante* ; 12e) *apretarse el cinturón* ; 13e) *importarle una cosa un carajo a alguien* ; 14e) *bajarse los pantalones* ; 15e) *guiñar un ojo* ; 16e) *echar la cremallera* 17e) *cerrar el pico* ; 18e) *cagarse en los pantalones*.

Afin de faciliter la compréhension des ef aux personnes qui n'ont pas une bonne compétence phraséologique en français et en espagnol, et pour ne pas avoir à transcrire le sens des expressions espagnoles dans cet article, nous avons recherché des ef équivalentes (par le sens, et par une composition lexicale la plus similaire possible) dans les deux langues et nous les avons désignées par un numéro équivalent. Ainsi, 1f) *adorer le veau d'or*, c'est-à-dire l'expression figée n° 1 française équivaut à l'expression figée n° 1 espagnole : 1e) *adorar el becerro de oro*.

Nous allons regrouper ces expressions dans des tableaux en fonction de leur niveau de langue, et nous indiquerons pour chacune d'entre elles l'information suivante :

- Compétence phraséologique des étudiants : nous allons traiter la compétence phraséologique des étudiants en analysant leur connaissance des EF au cours de la première et de la quatrième année de licence. Afin de contraster l'évolution de l'acquisition de cette compétence – toujours chez des locuteurs considérés comme natifs –, nous allons également l'analyser avec un groupe de personnes d'une cinquantaine d'années et de culture moyenne/supérieure. Chaque groupe est composé de 20 personnes. Les numéros qui apparaissent dans la deuxième et la troisième colonne après l'EF indiquent dans chaque cas le nombre de ces personnes qui connaissent (deuxième colonne) ou ne connaissent pas (troisième colonne) ces EF.

- Traitement de ces EF dans les dictionnaires monolingues et bilingues : nous allons observer la présence de ce groupe d'EF dans les dictionnaires monolingues, phraséologiques et bilingues les plus représentatifs de chaque langue. Dans le cas du français il s'agit du *Grand Robert* (GR) ; du *Trésor de la langue française informatisé* (TLFI), du *Dictionnaire des expressions et locutions d'A. Rey et S. Chantreau* (DEL) ; du *Gran Diccionario bilingüe Larousse « español-francés & francés-español »* (LBI) et du *Grand Espasa bilingüe « español-francés & francés-español »*. Dans le cas de l'espagnol, nous avons consulté le *Diccionario de uso del español* (DUE) ; le *Diccionario de la lengua española de la Real Academia* (DRAE) ; le *Diccionario fraseológico documentado del español actual* (DFDEA) ; le *Gran Diccionario bilingüe Larousse « español-francés & francés-español »* (LBI) et le *Grand Espasa bilingüe « español-francés & francés-español »*.
- Dans la colonne des définitions, nous utiliserons : a) une ou plusieurs définitions en fonction de chaque EF, en indiquant la source utilisée⁴, le registre de langue qui apparaît dans le dictionnaire ; b) l'EF équivalente lorsqu'elle apparaît dans le dictionnaire bilingue, c) le sens donné par les personnes consultées lorsque celui-ci s'éloigne du ou des sens possibles indiqués dans les ouvrages de référence consultés.

1.1 Registre cultivé, littéraire, vieilli

1.1.1 EF françaises

| Expressions françaises | EF connue | EF Non connue | GR | TLF | DEL | LBI | GEB | Définition(s) |
|---|--------------|---------------|----|--------------|-----|-----|-----|---|
| 1f) <i>Adorer le veau d'or</i> ⁵ | 3 8 12 | 17 12 8 | + | ⁵ | + | + | + | – avoir le culte de la richesse (GR). – <i>adorar el becerro de oro</i> (LBI). – <i>ambicionar, adorar las riquezas</i> (GEB). Agir bêtement. |
| 2f) <i>Franchir le Rubicon</i> | 0 0 13 | 0 20 7 | + | + | + | +1? | – | – Prendre une décision hasardeuse irrévocable et lourde de conséquences (TLF) – coup de force destiné à imposer qq chose (GR). – <i>pasar el Rubicón</i> (LBI). |

⁴ Dans notre base de données nous voulons cependant indiquer que chaque entrée équivaut à une seule définition.

⁵ Dans le TLF nous trouvons seulement *le veau d'or*.

| | | | | | | | | |
|--|----------------|----------------|------------|------------|------------|--------------|------------|--|
| 3f) <i>Sendormir sur ses lauriers</i> | 15 20 20 | 5 0 0 | + - | + + | - + | +1? - | + - | - jouir d'un repos mérité par de glorieux succès (GR). - Se contenter du premier succès (TLF). - dormirse en los laureles (LBI) / (GEB). Penser être tiré d'affaires. |
| 4f) <i>Tuer la poule aux œufs d'or</i> | 19 20 20 | 1 0 20 | + | + | + | + | - | - détruire par avidité ou impatience, la source d'un profit* important (GR). - <i>la gallina de los huevos de oro</i> ⁶ ! |
| 5f) <i>Se croire sorti de la cuisse de Jupiter</i> | 5 10 14 | 15 10 6 | + | + | + | + | + | - se croire de très haute naissance, et, par ext. , être très orgueilleux (GR). - <i>creerse descendiente de la pata del Cid, ser muy orgulloso</i> (LBI). - <i>creerse el ombligo</i> ou <i>el centro del mundo</i> (GEB). Faire le malin. Être noble. Être parfait. Être intelligent. |
| 6f). <i>Prendre la clé des champs</i> | 8 15 19 | 12 5 1 | - | - | + | + | - | - S'enfuir (DEL). - <i>tomar las de Villadiego</i> (LBI). |
| | | | 6/8 | 5/8 | 6/8 | 6/8 | 3/8 | |
| Expressions françaises | EF connues | EF non connues | GR | TLF | DEL | LBI | GEB | Définition(s) |
| 1f) <i>Adorer le veau d'or</i> | 3 8 12 | 17 12 8 | + | 7 | + | + | + | - avoir le culte de la richesse (GR) - <i>adorar el becerro de oro</i> (LBI) - <i>ambicionar, adorar las riquezas</i> (GEB) agir bêtement |
| 2f) <i>Franchir le Rubicon</i> | 0 0 13 | 0 20 7 | + + | + - | + - | +1? | - | - prendre une décision hasardeuse irrévocable et lourde de conséquences (TLF) - coup de force destiné à imposer qqch (GR) - <i>pasar el Rubicón</i> (LBI) |

6 Dans le LBI, nous trouvons seulement *la poule aux œufs d'or* dans l'entrée *poule* et nous trouvons cependant *matar la gallina de los huevos de oro* avec comme équivalent *tuer la poule aux œufs d'or* dans la partie espagnole. Dans le GEB nous ne trouvons que *la poule aux œufs d'or* dans l'entrée *poule*.

| | | | | | | | | |
|--|----------------|----------------|-----|-----|-----|-----|-----|---|
| 2e) <i>Pasar el Rubicón</i> | 0 0 4 | 20 20 16 | + | + | + | + | - | - <i>dar un paso decisivo después del cual ya no se puede retroceder</i> (DUE) - franchir le Rubicon (LBI) |
| 3e) <i>Dormirse en los laureles</i> | 13 20 20 | 7 0 20 | + | + | + | + | + | - <i>Cesar en un esfuerzo después de haber conseguido un triunfo</i> (DUE) - s'endormir sur ses lauriers (LBI)/(GEB) <i>no hay que confiarse, relajarse</i> |
| 4e) <i>Matar la gallina de los huevos de oro</i> | 13 18 20 | 7 2 0 | + | - | + | + | - | - <i>forzar la ganancia en una cosa, y destruirse la fuente de esa ganancia</i> (DUE) - tuer la poule aux oeufs d'or (LBI) |
| 5e) <i>Descender de la pata del Cid</i> | 4 8 8 | 16 12 12 | - | - | + | + | - | - <i>ser muy orgulloso</i> (DTemático) - se croire sorti de la cuisse de Jupiter (LBI) <i>creerse alguien importante</i> <i>creerse libre</i> |
| 6e) <i>Tomar las de Villadiego</i> | 5 10 17 | 15 10 3 | + | + | + | + | - | - <i>marcharse de un sitio precipitadamente o huyendo</i> (DUE) - prendre la clef des champs ou la poudre d'escampette (LBI) <i>cabrearse con alguien</i> |
| | | | 4/6 | 3/6 | 6/6 | 5/6 | 1/6 | |

Tableau 2

1.2 EF appartenant à un registre de langue de type standard, courant.

1.2.1 EF françaises

| Expressions françaises | EF connues | EF non connues | GR | TLF | DEL | LBI | GEB | Définition(s) |
|---------------------------------|----------------|----------------|----|-----|-----|-----|-----|---|
| 7f) <i>Être dans les nuages</i> | 20 20 20 | 0 0 0 | + | + | + | + | + | - être rêveur, distrait (TLF) - <i>estar en las nubes o en Babia</i> (LBI) - <i>estar en las nubes</i> (GEB) avoir l'esprit ailleurs rêvasser |

| | | | | | | | | |
|--|----|----|-----|-----|-----|-----|----------------|---|
| 8f) <i>Arrondir les angles</i> | 08 | 12 | + | + | - | + | + | - faire des concessions (GR) - <i>limar las asperezas</i> (LBI) - <i>limar asperezas</i> (GEB) faire qqch, mais pas dans les détails X parler avec nuance X |
| | 15 | 5 | | | | | | |
| | 20 | 0 | | | | | | |
| 9f) <i>Mettre la charrue avant les bœufs</i> | 15 | 5 | + | + | + | + | + | - faire d'abord ce qui devrait être fait ensuite, après (GR) - <i>empezar la casa por el tejado</i> (LBI) - <i>empezar la casa por la ventana</i> (GEB) aller trop vite pour faire qqch Brûler les étapes |
| | 20 | 0 | | | | | | |
| | 20 | 0 | | | | | | |
| 10f) <i>Tirer les marrons du feu</i> | 12 | 8 | + | + | + | + | - | - s'exposer à des risques... pour le seul profit d'autrui (GR) - <i>sacar las castañas del fuego</i> (LBI) Tirer des bénéfices |
| | 15 | 5 | | | | | | |
| | 17 | 3 | | | | | | |
| 11f) <i>Avoir une mémoire d'éléphant</i> | 20 | 0 | ? | + | + | - | - | - avoir une mémoire exceptionnelle et en particulier <u>être rancunier</u> ⁸ (TLF) - mémoire d'éléphant (GEB) avoir une grande mémoire |
| | 20 | 0 | | | | | | |
| | 20 | 0 | | | | | | |
| 12f) <i>Se serrer la ceinture</i> | 14 | 6 | - | + | + | 1 ? | - ⁹ | - ne pas manger à sa faim (TLF) - être privé du nécessaire (TLF) - se passer de qqch (GR) - <i>apretarse el cinturón</i> (LBI) |
| | 20 | 0 | - | + | - | - | - | |
| | 20 | 0 | + | - | + | - | - | |
| | | | 5/8 | 7/8 | 6/8 | 5/8 | 4/8 | |

Tableau 3

8 La plupart de personnes consultées n'utilisaient pas l'EF avec cette nuance.

9 Nous ne trouvons pas l'EF *se serrer la ceinture* sous *ceinture* mais par contre dans la partie espagnole sous *cinturón* nous trouvons *apretarse el cinturón* avec, comme équivalent, *se serrer la ceinture* !

1.2.2 EF espagnoles

| Expressions espagnoles | EF connues | EF non connues | DUE | DRAE | DFDEA | LBI | GEB | Définition(s) |
|--|----------------|----------------|--------|--------|--------|---------|--------|---|
| 7e) <i>Estar en las nubes</i> | 20 20 20 | 0 0 0 | + + | + - | + - | 1? - | + - | – <i>estar distraído</i> (DUE / GEB) – <i>ignorante de cierta cosa notoria</i> (DUE) – être dans les nuages (LBI / GEB) <i>estar en Babia ausente despiestado</i> |
| 8e) <i>Limar (las) asperezas</i> | 20 20 20 | 0 0 0 | + + | + + | + + | + + | + + | – <i>atenuar las diferencias entre personas</i> (DUE) – arrondir les angles (LBI / GEB) <i>olvidar el rencor olvidar los problemas</i> |
| 9e) <i>Empezar la casa por el tejado</i> | 15 18 20 | 5 2 0 | + + | - - | + + | + + | + + | – <i>empezar algo por donde debiera terminarse</i> (DUE) – mettre la charrue avant les bœufs (LBI / GEB) <i>Querer hacer las cosas muy rápidas</i> xx <i>Adelantarse a los acontecimientos.</i> |
| 10e) <i>Sacar(le) las castañas del fuego a alguien</i> | 14 20 20 | 6 0 0 | + + | + + | + + | + + | + + | – <i>hacerle algo que le corresponde hacer a él mismo, para sacarle de un apuro</i> (DUE) – tirer les marrons du feu (LBI / GEB) <i>querer hacer las cosas muy rápidas salir de un apuro ayudar</i> |

| | | | | | | | | |
|---------------------------------------|----|---|-----------------|-----|-----|-----|-----|---|
| 11e) <i>Tener memoria de elefante</i> | 16 | 5 | i ¹⁰ | - | i | - | - | – <i>tener mucha memoria.</i> |
| | 20 | 0 | | | | | | |
| | 20 | 0 | | | | | | |
| 12e) <i>Apretarse el cinturón</i> | 20 | 0 | + | + | + | + | + | – <i>tener que reducir por escasez de medios los gastos</i> (DRAE) – <i>se serrer la ceinture</i> (LBI / GEB) <i>gastar menos por no tener dinero</i> |
| | 20 | 0 | | | | | | |
| | 20 | 0 | | | | | | |
| | | | 7/8 | 4/8 | 5/8 | 5/8 | 5/8 | |

Tableau 4

1.3 EF appartenant à un registre de langue considéré comme (vulgaire, populaire, colloquial, argotique)

1.3.1 EF françaises

| Expressions françaises | EF connues | EF Non connues | GR | TLF | DEL | LBI | GEB | Définition(s) |
|---|------------|----------------|----|-----|-----|-----|-----|---|
| 13f) <i>N'en avoir rien à foutre de</i> | 20 | 0 | + | + | 1 ? | - | + | – <i>n'avoir pas à tenir compte de</i> (GR) (vulg) – <i>se moquer de</i> (GR) (vulg) – <i>no pintar nada en</i> (GEB) ? – <i>ni importarle nada</i> (GEB) ne pas être intéressé par |
| | 20 | 0 | | | | | | |
| | 20 | 0 | + | - | - | - | - | |
| 14f) <i>Baisser son froc</i> | 14 | 6 | + | + | + | - | + | subir une humiliation sans protester, se soumettre (GR) (fam.) – <i>bajarse los pantalones</i> (GEB) se faire avoir enlever son pantalon être d'accord pour se laisser avoir |
| | 15 | 5 | | | | | | |
| | 20 | 0 | | | | | | |

10 Nous trouvons dans le DUE, et dans le DFDEA *memoria de elefante*.

| | | | | | | | | |
|---------------------------------|----------------|---------------|-----|-----------------|-----------------|-----|-----------------|---|
| 15f) <i>Faire de l'œil</i> | 15 18 20 | 5 2 0 | + | + | - | 1 ? | 1 ? | - cligner de l'œil pour attirer son attention, le provoquer ou pour marquer qu'on est de connivence avec lui (GR) (fam.) - lui adresser des regards amoureux / DEL) - <i>guiñar</i> (LBI / GEB) draguer |
| 16f) <i>Fermer sa boîte</i> | 8 10 16 | 12 10 4 | + | + | + | - | - | - se taire (TLF)(fam.) être mort |
| 17f) <i>Fermer son bec</i> | 18 20 20 | 8 0 20 | - | + | - | + | - ¹¹ | se taire (TLF)(fam.) |
| 18f) <i>Faire dans son froc</i> | 20 20 20 | 0 0 0 | + | + ¹² | + ¹³ | - | + | - avoir très peur (GR) (fam.) - <i>hacérselo encima</i> (GEB) |
| | | | 6/8 | 6/8 | 4/8 | 2/8 | 3/8 | |

Tableau 5

1.3.2 EF espagnoles

| Loc. verbales esp. | EF connues | EF non connues | DUE | DRAE | DF DEA | LBI | GEB | Définition(s) |
|---|----------------|----------------|-----|------|--------|-----|-----|---|
| 13e) <i>Importarle una cosa un carajo a alguie.</i> | 20 20 20 | 0 0 0 | + | + | + | + | - | - n'en avoir rien à foutre (LBI) <i>ni importar nada una cosa</i> XXXXX |
| 14e) <i>Bajarse los pantalones</i> | 20 20 20 | 0 0 0 | + | + | + | + | + | - <i>ceder en condiciones poco honrosas</i> (RAE) (vulg.) - baisser sa culotte, son pantalon (LBI / GEB) |

11 *Fermer son bec* n'apparaît pas dans la partie française, mais son équivalent *cerrar el pico* apparaît dans la partie espagnole où on trouve aussi *fermer son bec*.

12 Le TLF signale *faire dans son froc* comme populaire.

13 Le dictionnaire des locutions le signale comme argotique.

| | | | | | | | | |
|---|----|----|-----|-----|-----|-----|-----|---|
| 15e) <i>Guiñar un ojo</i> | 20 | 0 | + | - | - | - | - | – cerrar un poco y abrir a continuación uno de ellos o los dos, por enfermedad o para hacer señas (DUE) |
| | 20 | 0 | | | | | | |
| | 20 | 0 | | | | | | |
| 16e) <i>Echar la cremallera</i> | 10 | 10 | + | - | - | - | - | – dejar de hablar (DUE) |
| | 15 | 5 | | | | | | |
| | 15 | 5 | | | | | | |
| 17e) <i>Cerrar el pico</i> | 20 | 0 | + | + | + | 1? | + | – callar. (DRAE) (col.) – no decir algo que se sabe (DUE) fermer son bec, la fermer (LBI / GEB) |
| | 20 | 0 | + | - | - | - | - | |
| | 20 | 0 | | | | | | |
| 18e) <i>Cagarse (en) los pantalones</i> | 20 | 0 | - | - | + | - | - | – ser dominado por el miedo (DFDEA) (vul.) |
| | 20 | 0 | | | | | | |
| | 20 | 0 | | | | | | |
| | | | 6/7 | 3/7 | 4/7 | 3/7 | 2/7 | |

Tableau 6

Nous allons réaliser, pour suivre, une analyse globale des deux ensembles de données traités dans cette étude. Nous allons les illustrer à l'aide de plusieurs tableaux résumés qui incorporent toutes les données, à savoir, la compétence phraséologique par tranche d'âge et le traitement de ces EF par les dictionnaires analysés.

2 Compétences phraséologiques par tranche d'âge pour les EF françaises et les EF espagnoles

Nous avons indiqué dans les divisions précédentes les chiffres qui correspondent à la compétence phraséologique des personnes consultées pour chaque EF par tranche d'âge. Ces nombres nous permettent tout d'abord de réaliser une comparaison, cas par cas, en utilisant les chiffres qui figurent devant chaque EF dans les deux langues.

Nous montrons ci-dessous les chiffres globaux de la compétence phraséologique des trois groupes de sujets analysés, en fonction des trois niveaux de langue des EF.

| | Groupe F : 18 ans | Groupe F : 22 ans | Groupe F : 50 ans | Groupe E : 18 ans | Groupe E : 22 ans | Groupe E : 50 ans |
|-------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| E1 (vieilli/littéraire) | 50 = 41,6% | 73 = 60,8% | 98 = 81,6% | 35 = 25,1% | 66 = 55% | 76 = 63,3% |
| E2 (standard) | 89 = 74% | 110 = 91,6% | 117 = 97,5% | 108 = 90% | 116 = 96,6% | 120 = 100% |
| E3 (populaire/familier) | 100 = 78% | 110 = 77% | 116 = 96,6% | 110 = 91,6% | 115 = 95,8% | 115 = 95,8% |

Tableau 7

2.1 Compétences phraséologiques réelles

Dans les colonnes des définitions, nous avons inséré à la suite de celles-ci, le sens que leur donnaient certains des sujets questionnés. Étant donné que beaucoup de ces définitions sont incorrectes, nous avons dû recalculer les compétences en fonction de ces erreurs.

| | Groupe F : 18 ans | Groupe F : 22 ans | Groupe F : 50 ans | Groupe E : 18 ans | Groupe E : 22 ans | Groupe E : 50 ans |
|-------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| E1 (vieilli/littéraire) | 38 = 31% | 65 = 54,1% | 86 = 71% | 31 = 26% | 60 = 50% | 70 = 58,3% |
| E2 (standard) | 83 = 70% | 99 = 82% | 109 = 90,8% | 98 = 81,6% | 106 = 91,6% | 115 = 95,8% |
| E3 (populaire/familier) | 95 = 80% | 93 = 77,5% | 116 = 96,6% | 103 = 85,3% | 111 = 92,5% | 115 = 95,8% |

Tableau 8

Les résultats réels nous permettent de réaliser une comparaison globale afin de comparer la compétence phraséologique des sujets analysés dans les deux langues. Nous observons alors dans les deux langues analysées que :

- la compétence phraséologique est toujours inférieure dans les trois niveaux de langue avec les sujets qui ont 18 ans ;
- la compétence augmente considérablement entre 18 ans et 22 ans (il faudrait dans ce cas compléter ces recherches avec des locuteurs qui n'aient pas réalisé d'études universitaires afin de contraster les résultats) ;
- la compétence des trois groupes est beaucoup plus grande avec les EF qui appartiennent aux niveaux de langue standard et populaire/familier.

La compétence est nettement inférieure dans les deux langues avec les expressions qui appartiennent au registre de langue cultivé/formel, vieilli/littéraire.

Les résultats sont équivalents dans les deux langues.

3 Traitement des EF dans les dictionnaires

3.1 Dictionnaires français

| | GR | TLF | DEL | LBI | GEB | |
|-----------------------------------|---------------|-------------|---------------|-------------|-------------|-------------|
| EF cultivé/formel, litt., vieilli | 6/8 | 5/8 | 6/8 | 6/8 | 3/8 | 23/32 = 71% |
| Registre standard | 5/8 | 7/8 | 6/8 | 5/8 | 4/8 | 23/32 = 71% |
| Registre populaire/familier | 6/8 | 6/8 | 4/8 | 2/8 | 3/8 | 18/32 = 56% |
| | 19/24 = 79,1% | 18/24 = 75% | 16/24 = 66,6% | 13/24 = 54% | 10/24 = 41% | |

Tableau 9

3.2 Dictionnaires espagnols

| | DUE | DRAE | DFDEA | LBI | GEB | |
|-----------------------------------|---------------|--------------|--------------|------------|------------|-------------|
| EF cultivé/formel, litt., vieilli | 4/6 | 3/6 | 6/6 | 5/6 | 1/6 | 18/24 = 75% |
| Registre standard | 7/8 | 4/8 | 5/8 | 5/8 | 5/8 | 21/24 = 87% |
| Registre populaire/familier | 6/7 | 3/7 | 4/7 | 3/7 | 2/7 | 16/24=66% |
| | 17/21 = 70,8% | 10/21= 47,6% | 15/21= 71,4% | 13/21= 61% | 8/21 = 38% | |

Tableau 10

L'analyse du traitement des formes contenues dans les dictionnaires nous montre :

- que les dictionnaires n'incorporent pas de nombreuses EF. Nous tenons à rappeler que nous avons choisi des EF très courantes dans tous les cas ;
- que les dictionnaires monolingues et bilingues enregistrent davantage d'EF appartenant aux registres de langue 1 et 2 et que le nombre d'expressions appartenant au registre populaire/familier est inférieur dans les dictionnaires monolingues et très inférieur dans les dictionnaires bilingues. Nous utilisons comme preuve à l'appui les pourcentages qui apparaissent dans la colonne inférieure des tableaux 9 et 10 ;
- que le nombre d'EF présentes dans les dictionnaires bilingues est nettement inférieur à celui présent dans les dictionnaires monolingues. Ceci pose un double problème car mis à part l'absence de ces formes, les traducteurs doivent faire un travail de recherche supplémentaire dans d'autres ouvrages spécialisés monolingues pour trouver le sens et l'équivalent.

Les chiffres d'inclusion des EF dans les dictionnaires coïncident avec ceux d'une étude plus poussée que nous sommes en train de réaliser (Mogorrón 2006, 2008, 2009).

Dans ces articles, nous avons analysé la présence discrète des constructions verbales figées (CVF) dans les dictionnaires, à partir d'une base de données de ces CVF que le groupe FRASYTRAM est en train de constituer. Ainsi, pour un total de 15 913 CVF, retenues, sélectionnées et classées, nous avons trouvé les chiffres suivants¹⁴ : DEA 363, DRDP 71, DTDFH 855, Dtemático 166, DFDEA, 609, EPM, 149, Espasa Calpe 1139, LARBI 298, LIBSA 126, DUE 6478, RAE 2612, sans référence dans les dictionnaires : 2152, etc. (figure 1).

14 Le procédé suivi est toujours le même : nous vérifions la présence de ces CVF dans les dictionnaires en suivant le même ordre : 1) DUE ; 2) DRAE ; 3) DTDFH ; 4) DFDEA, ; 5) DEA ; 6) LARBI, etc.

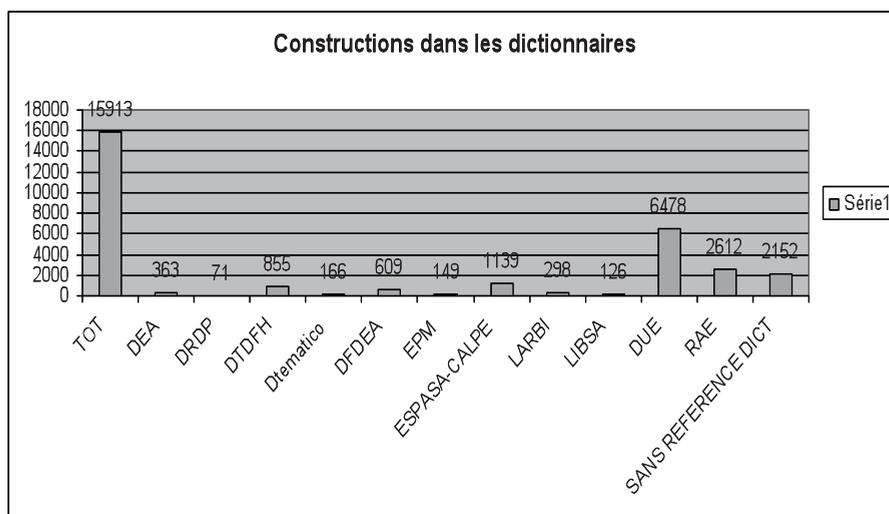


Figure 1

D'un côté, cela signifie que 9090 CVF apparaissent dans les deux dictionnaires les plus utilisés (DRAE et DUE). Mais d'un autre côté, nous pouvons également interpréter que 6823 CVF c'est-à-dire 43% des séquences retenues n'y apparaissent pas et même que 2612 CVFS c'est-à-dire 16,5% du total n'apparaissent dans aucun dictionnaire.

4 EF polysémiques

Nous voyons également que de nombreuses EF du petit corpus traité relèvent de la polysémie :

- en français les expressions : 2f) *franchir le Rubicon* ; 3f) *s'endormir sur ses lauriers* ; 12f) *se serrer la ceinture* ; 15f) *faire de l'œil* ;
- en espagnol : 7e) *estar en las nubes* ; 17e) *cerrar el pico*.

Mais le problème principal pour la traduction n'est autre que ces EF sont polysémiques dans une langue et pas dans l'autre, comme le montre le tableau suivant :

| EF françaises | Définition(s) | EF dans le dict. bilingue | EF espagnoles | Définition(s) | EF dans le dict. bilingue |
|--|--|--|---|--|-------------------------------------|
| 2f) <i>franchir le Rubicon</i> | <ul style="list-style-type: none"> – prendre une décision hasardeuse irrévocable et lourde de conséquences (TLF) – coup de force destiné à imposer qqch (GR) | <i>pasar el Rubicón</i> (LBI) | 2 ^e) <i>pasar el Rubicón</i> | – <i>Dar un paso decisivo después del cual ya no se puede retroceder</i> (DUE) | franchir le Rubicon (LBI) |
| 3f) <i>s'endormir sur ses lauriers</i> | <ul style="list-style-type: none"> – jouir d'un repos mérité par de glorieux succès (GR) – se contenter du premier succès (TLF) | <i>dormirse en los laureles</i> (LBI) | 3 ^e) <i>dormirse en los laureles.</i> | <i>Cesar en un esfuerzo después de haber conseguido un triunfo</i> (DUE) | – s'endormir sur ses lauriers (LBI) |
| 7f) <i>être dans les nuages</i> | – être rêveur, distrait (TLF) | – <i>estar en las nubes o en Babia</i> (LBI) | 7e) <i>estar en las nubes.</i> | <ul style="list-style-type: none"> – <i>estar distraído.</i> (DUE) – <i>ignorante de cierta cosa notoria</i> (DUE) | être dans les nuages (LBI) |
| 12f) <i>se serrer la ceinture</i> | <ul style="list-style-type: none"> – ne pas manger à sa faim (TLF) – être privé du nécessaire (TLF) – se passer de qqch (GR) | – <i>apretarse el cinturón</i> (LBI) | <i>apretarse el cinturón.</i> | <i>tener que reducir por escasez de medios los gastos.</i> (DRAE) | se serrer la ceinture (LBI) |
| 15f) <i>faire de l'œil</i> | <ul style="list-style-type: none"> Cligner de l'œil pour attirer son attention, le provoquer ou pour marquer qu'on est de connivence avec lui (GR) (fam.). – Lui adresser des regards amoureux /DEL) | – <i>guiñar</i> (LBI) | 15 ^e) <i>Guiñar un ojo.</i> | <i>Cerrar un poco y abrir a continuación uno de ellos o los dos, por enfermedad o para hacer señas</i> (DUE) | |
| 17f) <i>fermer son bec</i> | Se taire (TLF) (fam.) | | 17e) <i>cerrar el pico.</i> | <ul style="list-style-type: none"> – <i>callar.</i> (DRAE) (col.) – <i>no decir algo que se sabe</i> (DUE) | fermer son bec, la fermer (LBI) |

Tableau 11

Nous observons dans les colonnes 3 et 6 du tableau 11 que les dictionnaires bilingues dans chaque cas (2, 3, 7, 12, 15, 17) ne donnent qu'une EF dans l'autre langue ce qui pose des problèmes car à quel sens se réfèrent-ils ? Cet aspect n'est donc pas traité rigoureusement dans les dictionnaires.

5 Équivalences de sens et paronymie

Le premier problème que nous avons observé lors de l'analyse des EF dans les dictionnaires bilingues, est que, parfois, il n'y a pas d'homogénéisation des incorporations dans les deux parties espagnol-français/français-espagnol. Nous avons déjà signalé dans les notes 6, 10 et 12 comment certaines EF ne sont pas incluses dans la partie français-espagnol du LBI (*se serrer la ceinture, fermer son bec*) et comment par contre elles apparaissent comme équivalents d'EF espagnoles dans la partie *español-francés*.

L'analyse des équivalents donnés nous a également permis de trouver une erreur dans le tableau 3. Le GEB donne comme équivalent de *mettre la charrue avant les bœufs* (voir tableau 3) *empezar la casa por la ventana* qui n'existe pas en espagnol. Dans la partie *español-francés* nous trouvons à *casa* : *echar o tirar la casa por la ventana* et *empezar la casa por el tejado*. Nous supposons que les lexicographes ont mélangé ces deux EF involontairement.

Il y aurait également d'autres problèmes qui ne sont pas traités en profondeur dans les dictionnaires monolingues et bilingues tels que, par exemple, les équivalences de sens. Traditionnellement la classification utilisée par la plupart des dictionnaires consiste à suivre un critère d'ordre alphabétique des unités appelé « sémasiologique ». Cependant, de nombreuses expressions sont paronymes, mais très peu de dictionnaires offrent ce genre d'équivalents et réalisent une classification onomasiologique. Nous avons ainsi certains ouvrages qui, bien qu'ils suivent cette classification sémasiologique, offrent certains équivalents : le DUE en espagnol, le *Grand Robert* en français, et d'autres dictionnaires qui comme le *Diccionario Temático de Locuciones Francesas con su correspondencia española*, le *Diccionario Temático de frases hechas*, *Le Bouquet des expressions imagées*, etc. réalisent cette classification en partant de l'idée et du sens.

Nous avons demandé aux groupes de personnes consultées antérieurement quelles expressions elles connaissaient pour exprimer le sens de « être distraité, être dans la lune » ? Les résultats sont :

- pour le français : *avoir la tête ailleurs, être dans les nuages, être dans la lune, être à l'ouest, être sur Mars, être tête en l'air.*
- pour l'espagnol : *estar en las nubes, no estar en lo que se está, irse por los cerros de Úbeda, estar apavao, estar en la inopia, no saber dónde se tiene la cabeza, no estar al loro.*

Nous voyons que les personnes consultées connaissent fréquemment des EF parasyonymes, mais généralement elles n'en connaissent qu'une petite partie. Pour essayer de pallier ces carences et ces difficultés de compréhension et de traduction et afin que les utilisateurs puissent connaître ces unités et leur sens, le groupe de recherche FRASYTRAM de l'Université d'Alicante est en train de créer un outil d'une grande utilité pour les traducteurs et utilisateurs de la langue. À cette fin, nous sommes en train d'élaborer une liste, la plus exhaustive possible, de constructions verbales dans le but d'essayer de corriger ces carences et ces difficultés de compréhension et de traduction. Nous avons ainsi dans les cas des EF parasyonymes de *être dans la lune* :

- en espagnol : *estar bailando en belén* (RAE), *estar en la luna* (DUE), *estar en las nubes* (DUE), *estar en el limbo*, *estar en la higuera*, *estar en la parra*, *estar en las Batuecas* (DUE), *estar pensando en la mona de Pascua* (DTDFH), *tener la cabeza a las once* (RAE), etc.
- en français : *avoir la tête dans les nuages* (CD), *être dans la lune* (GR), *être, vivre dans les nuages* (TLF), *être dans les limbes* (GR), *n'avoir pas les pieds sur terre* (GR), *bayer aux corneilles* (TLF), etc.
- en italien : *essere fra le nuvole*, *vivere fra le nuvole*, *avere il capo / la testa fra le nuvole*, *vivere la testa nelle / tra le nuvole*, *vivere nel mondo della luna*, *vivere nel mondo dei sogni*, etc. ;
- en russe : *витать в облаках*, *жить в облаках*, *витать в эмпиреях*, *носиться в эмпиреях*, *парить в эмпиреях*, *строить воздушные замки*, etc. ;
- en anglais : *to have one's head in the clouds*, *to be miles Hawaii*, *to be spacey*, *to be in the clouds*, *to be lost in thought*, *to be wrapped in thought*, *to be lost in (a) reverie*, etc. ;
- en catalan-valencien : *estar en la lluna*, *estar en la lluna de València*, *estar en la parra*, *estar empanat*, *estar pensant en les mussaranyes*, *estar al cel de les oques*, *estar a tres quarts de quinze*, *estar en el tercer cel*, *pensar en la mona de Pasqua*, etc.

Conclusion

Il existe quantité d'expressions figées dans les systèmes phraséologiques de toutes les langues, mais le traitement de ces EF dans les dictionnaires monolingues et bilingues est insuffisant dans de nombreux cas. Nous avons ici observé que la compétence phraséologique des usagers est moins forte avec les EF appartenant au niveau cultivé, littéraire, tandis que les dictionnaires incluent ces EF cultivées et littéraires beaucoup plus fréquemment que celles qui appartiennent au registre de langue familier/populaire et qui sont cependant beaucoup plus utilisées par les usagers des langues. La traduction de ces formes n'a pas été non plus traitée exhaustivement par les dictionnaires bilingues dans lesquels les problèmes de la polysémie et de la parasyonymie ne sont pas traités exhaustivement. Il faudrait

donc créer de grandes bases de données qui permettent aux usagers d'une langue de pouvoir consulter toutes ces EF afin d'en connaître le sens ainsi que les équivalents possibles afin de pouvoir essayer de les traduire ou d'en trouver le meilleur équivalent possible.

Bibliographie

- Alvar Ezquerro (M.), 1981 : « Los diccionarios bilingües: su contenido », dans *Lingüística española actual*, II:1 ; 175-196 (reeditado en el capítulo 8 de *Lexicografía descriptiva*, 1993, Barcelona, bibliograf).
- Alvar Ezquerro, (M.), 2002 : *Los diccionarios de ayer y hoy*, Granada, Comares.
- Blanco Escoda, (X.), 1999 : « ¿Cómo leer un diccionario? Algunos aportes de la lingüística informática », dans Fróis, (J.), Barbas (M.), éd., *Cenários interactivos. Arte e tecnologia*. Santarém, E.S.E. de Santarém, p. 117-127.
- Blanco Escoda (X.), 2002 : « Les déterminants figés », dans *Langages*, n° 145, p. 61-81.
- Corpas Pastor (G.), 1997 : *Manual de fraseología española*, Madrid, Gredos.
- Gross (M.), 1982 : « Une classification des phrases figées du français », dans *Revue québécoise de linguistique*, 11/2, p. 151-185.
- González Rey (I.), 2002 : *La phraséologie du français*, Toulouse, Presses Universitaires du Mirail.
- Martin (R.), 1996 : *Sur les facteurs du figement lexical. La locution : entre langue et usage*, Paris, Fontenay/Saint-Cloud.
- Mejri (S.), 1997 : *Le figement lexical. Descriptions linguistiques et structuration sémantique*, Tunis, Publication de la faculté des lettres de la Manouba.
- Mejri (S.), 2000 : « Figement et dénomination », dans *Meta*, 45, p. 609-621.
- Mogorrón Huerta (P.), 2002 : *La expresividad en las locuciones verbales en francés y en español*. Alicante, Publicaciones de la Universidad de Alicante.
- Mogorrón Huerta (P.), 2004 : « Los diccionarios electrónicos fraseológicos, perspectivas para la lengua y la traducción », dans *Estudios de Lingüística*, 1, p. 381-400.
- Mogorrón Huerta (P.), 2005 : « As locuciones verbais nos dicionarios », dans *Cadernos de Fraseoloxía Galega*, 7, p. 149-176.
- Mogorrón Huerta (P.), 2008 : « Compréhension et traduction des locutions verbales », dans *Meta*, 53/2, p. 378/406.
- Ruiz Gurillo (L.), 1997 : *Aspectos de fraseología teórica española*, València, Universitat.
- Zuluaga Osþina (A.), 1980 : *Introducción al estudio de las expresiones fijas*, Frankfurt a. M., Lang.

Dictionnaires

- DEA = *Diccionario del español Actual* (2005³/1999). Madrid, Aguilar Lexicografía.
- DFDEA = *Diccionario fraseológico documentado del español actual* (2004). Madrid, Aguilar Lexicografía.
- DRAE = *Diccionario de la Real Academia Española* (1992). Madrid, Espasa Calpe.

- Dtemático = *Diccionario temático de locuciones francesas con su correspondencia española* (2004), de J. Sevilla Muñoz / J. Cantera de Urbina. Madrid, Gredos.
- DTDFH = *Diccionario Temático de Frases Hechas* (2004), de S. Rodríguez-Vida, Barcelona, Columbus.
- DUE = *Diccionario de uso del español* (1999/1977), de María Moliner, Madrid, Gredos.
- AR = *Dictionnaire des expressions et locutions figurées* (1979/+ index final 1982), de A. Rey/S. Chantreau, Paris, Larousse.
- GEB = *Gran Espasa bilingüe español-francés / français-espagnol*. (2000), Espasa, Madrid.
- Dictionnaire Larousse de la langue française*, Lexis. (1991). Larousse. Paris.
- GR = *Le Grand Robert* de la Langue française. Version 2 CD-Rom 2005/2006 Le Robert/Sejer. www.lerobert.com.
- LARBI = *Larousse moderno français-espagnol español-francés* (1993/1967), de R. García-Pelayo / J. Testas, Paris, Larousse.
- TLF = *Le Trésor de la Langue Française* informatisé. ISBN : 2-271-06273-X. www.cnrseditions.fr. www.tlfi.fr.

Abdelaziz Berkai,
Département de langue et culture amazighes,
Université de Béjaïa,
Béjaïa,
Algérie,
aberkai@yahoo.fr

Quelques problèmes lexicographiques que posent l'établissement d'équivalences et leur organisation dans un dictionnaire kabyle-français

Résumé

Nous nous proposons d'aborder ici le problème d'équivalence interlinguistique, kabyle-français en particulier, du point de vue de l'établissement des équivalences elles-mêmes et de leur classement macro- et microstructurel. Nous examinerons d'abord les types de relations d'équivalence envisageables en lexicographie bilingue en rapport avec la notion d'hétéronymie, et aborderons ensuite à travers quelques exemples le problème des locutions, de l'équivalence et du classement macrostructurel, après avoir abordé celui de l'ordre des acceptions des mots entrées en microstructure.

Mots-clés : équivalence interlinguistique ; hétéronymie ; lexiculture ; locution ; ordre (des acceptions).

Introduction

L'équivalent du bilingue est comparable au synonyme du monolingue : c'est un « synonyme interlingual » ou, comme l'appelle R. Jakobson, un « hétéronyme » (Landheer 1981 : 149). Mais, aussi bien dans la synonymie intralinguistique et plus encore dans l'hétéronymie interlinguistique, la correspondance ou l'interchangeabilité est rarement totale. L'équivalence totale, corollaire généralement de la monosémie des mots, n'existe réellement que dans certains secteurs spécialisés de la langue, comme l'électronique, l'informatique...

On a souvent affaire à une correspondance partielle et parfois même, pour certaines notions appelées *mots culturels* à la suite de Ladislav Zgusta, à l'absence d'équivalence dans l'autre langue. Un mot comme [1] *tingett* [θingətt], « mèche de cheveux que l'on coupe sur le front de la jeune mariée », relevé dans le parler kabyle d'Aokas, en est un. C'est que chaque langue, et c'est bien connu, constitue un découpage propre de la réalité, reflet d'une culture, d'une histoire, d'un vécu propres. Au découpage du français, par exemple, de la notion de « voie », selon qu'elle se situe ou

non dans l'agglomération, qu'elle soit plantée d'arbres ou non, qu'elle soit revêtue ou non, en : *rue* « voie bordée de maisons » ; *boulevard* « large voie faisant le tour d'une ville » ou « rue très large, généralement plantée d'arbres » ; *avenue* « voie plantée d'arbres qui conduit à une habitation », etc. ; *chemin* et *route* en dehors de l'agglomération ; *passage* « endroit par où l'on peut passer (allée, couloir, galerie...) » ; *piste* « chemin non revêtu » ; correspond en kabyle globalement une seule dénomination : *abrid*. Inversement, le kabyle (parler d'Aokas) distingue entre l'astre *soleil* appelé *tittict* [θitʰtʰijθ] et son rayonnement, sa lumière, sa chaleur, appelés *tafukt* [θafuçθ] et l'ensemble appelé *ittij* [itʰtʰiʒ], alors que le français se suffit de *soleil* pour exprimer l'ensemble. Pendant qu'une langue utilise un seul mot pour nommer une réalité, une autre recourt à une périphrase. Au mot kabyle [2] *qqunnez* correspond une périphrase en français : « s'avouer vaincu au jeu de devinettes ; donner sa langue au chat » (Dallet). Même dans le cas d'un isomorphisme sémantique, les mots ont toujours des charges culturelles et des symboliques propres aux communautés qui les utilisent que Robert Galisson appelle *lexiculture*, c'est-à-dire « la culture mobilisée et actualisée dans et par les mots » (Song 2003 : 375), « une valeur ajoutée que tout le monde connaît et qui pourtant est presque toujours absente des dictionnaires censés refléter la signification complète des mots » (Pruvošt 2005 : 18). Mais si cette tâche de donner des mots des « significations complètes » est réalisable, elle est d'abord celle du monolingue, sur lequel doit s'appuyer objectivement le bilingue et auquel il renvoie implicitement d'ailleurs.

L'absence de dictionnaires monolingues en berbère doit être compensée par l'enrichissement de l'information microstructurelle, notamment à travers des développements encyclopédiques et une exemplification fournie, particulièrement dans les cas d'absence totale d'équivalence. En l'occurrence, et dans la mesure du possible, un schéma, un dessin ou une image sont souvent plus parlants que les meilleures gloses. L'image d'un métier à tisser (*azeffa*), d'un moulin à eau berbère, d'une femme battant du lait assise à même le sol tenant son petit-fils dans son giron, d'une parure d'argent aux motifs berbères portée par une femme habillée de manière traditionnelle, etc., sont, plus que de simples illustrations, des supports à forte charge culturelle et mémorielle irremplaçables dans un dictionnaire. « La monographie encyclopédique », développement récent de la lexicographie bilingue, permet à l'usager « de mieux appréhender la réalité étrangère » (Duval 2003 : 73).

Nous nous proposons d'aborder dans ce qui suit le problème d'équivalence interlinguistique, en particulier kabyle-français, du point de vue de l'établissement des équivalences elles-mêmes et de leur classement macro- et microstructurel. Nous commencerons par définir les types de correspondance interlinguistique possibles dans un dictionnaire en les articulant autour de la notion d'hétéronymie.

1 Les différents types de correspondance interlinguistique

Trois types de relation d'équivalence sont envisageables dans un dictionnaire bilingue :

1.1 L'hétéronymie ou plus précisément la « monohétéronymie »

L'entrée E possède un seul équivalent Eq qui la traduit totalement (l'inverse n'étant pas toujours vrai). Trois cas peuvent se présenter en l'occurrence :

- a) E et Eq sont isomonosémiques : $E = Eq$. Cela concerne surtout les lexiques spécialisés.

[3] *astillu* ou *imru* (néologisme) : « stylo ».

- b) E et Eq sont isopolysémiques : $E (E_1, E_2, \dots, E_n) = Eq (Eq_1, Eq_2, \dots, Eq_n)$, avec : $E_1 = Eq_1, E_2 = Eq_2, \dots, E_n = Eq_n$. C'est le cas le moins fréquent.
- c) $E \subset Eq$. Le sémantisme de Eq est plus large que celui de E . C'est un cas qui n'est pas rare.

[4] *akured* : « puce (insecte parasite) ».

Le mot *puce* a en français au moins trois sens différents, alors qu'il n'en a qu'un seul en kabyle. Dans ce cas une indication sémantique ajoutée entre parenthèses est nécessaire pour préciser le sens requis.

1.2 La « polyhétéronymie »

E possède plusieurs équivalents partiels. $E = Eq_1; Eq_2, \dots, Eq_n$. C'est le cas le plus fréquent. Pour chaque équivalent partiel, généralement polysémique, il faudrait préciser (au besoin) le sens requis.

[5] *azru* « 1. caillou ; pierre ; rocher. 2. pile (électrique) ».

1.3 L'« anhétéronymie » ou l'absence d'équivalent dans la langue cible

L'entrée est un mot culturel traduisant une réalité concrète ou abstraite spécifique à la communauté sociolinguistique en question. Ainsi en va-t-il des mots qu'on trouve dans divers secteurs de la langue comme la gastronomie, les pratiques religieuses, institutionnelles, matrimoniales, festives, sexuelles, etc. En l'occurrence, il convient d'expliquer le mot par une définition, comme dans un dictionnaire monolingue, qui irait au besoin par un développement encyclopédique au-delà du sens strict du mot qui risque de ne rien vouloir dire pour l'utilisateur passif. Définir [6] *acerrid* [*afærrid*] comme « galette de pain faite avec de la farine de gland mélangée à de l'eau salée » est sémantiquement juste, mais ajouter que sa pâte est dure et qu'on la mangeait en Kabylie à une époque, maintenant révolue, où la population vivait dans la misère et qu'elle servait à suppléer la galette de semoule qui était alors un luxe qu'on ne pouvait pas s'offrir tout le temps, est mieux. Cette idée de misère et de faim à laquelle renvoie ce mot pour l'ensemble de ses usagers, ceux de la troisième et de la quatrième génération notamment, n'est pourtant pas contenue dans

sa définition minimale entre guillemets. Ce type d'information, qu'on peut appeler avec Robert Galisson (1987 : 119) « charge culturelle partagée (CCP) », permet le passage de la langue à la culture qui lui est « consubstantielle », donc nécessaire à sa bonne appréhension.

L'absence d'équivalence peut aussi concerner une réalité qui existe, mais qui n'est pas nommée dans la langue cible. C'est ce que nous pouvons appeler, faute de mieux, « désignation culturelle ». Dans ce cas, l'équivalence est facile à établir. Elle peut s'établir au mieux par un mot composé, au pire par une glose. Un mot kabyle (parler d'Aokas) comme [7] *ibunda* (sing. *abandu*¹) [*ivunda*], « petites excroissances ayant la forme de grains de blé qui se forment sur le bout des seins chez la femme enceinte » (Rahmani 1934 : 20), exprime bien une réalité universelle que la langue cible n'a semble-t-il pas dénommée. [8] *Aqennuz* « aliments mâchés que l'on tient dans la bouche et qui font gonfler les joues ; objet dérobé que le voleur essaye de cacher sous son habit, mais qui fait saillie » (Huyghe 1904 : 20), en est un autre.

2 L'ordre des acceptions

L'équivalence est plus souvent partielle et concerne surtout des unités polysémiques qui posent pour le lexicographe le problème de l'ordre des acceptions dans l'article. Cet ordre doit reposer sur des critères préalablement établis et s'appliquer de façon cohérente à l'ensemble des unités concernées. On peut choisir entre des critères logiques, fréquentiels, distributionnels et historiques « sans s'enfermer, bien entendu, dans une procédure de sélection exclusive » (Szende 2000 : 70). L'ordre logique qui permet d'aller du sens général, commun, au sens particulier, technique ; du sens propre au sens figuré, par extension, etc., paraît le mieux indiqué, parce que plus pratique. Mais, quand la fréquence est avérée, mesurée, concernant les acceptions d'une unité, l'ordre fréquentiel s'impose parce qu'il concerne l'usage : le sens le plus usuel doit être traité en premier lieu. Mais « si la fréquence de tel ou tel mot peut être facilement vérifiée à l'aide de logiciels aujourd'hui courants, il n'existe pas de données précises concernant la fréquence d'emploi des différentes valeurs et acceptions des mots » (Szende 2000 : 70). Cette fréquence repose en réalité, comme le remarque Bernard Al, cité par Szende (*ibid.*), sur l'idée que s'en font les lexicographes eux-mêmes. L'absence de grands corpus textuels informatisés pour le berbère complique encore davantage l'opération. Pour le dictionnaire kabyle-français l'ordre logique s'impose donc de lui-même et c'est du reste celui qui est généralement adopté dans les dictionnaires, même dans ceux de langues à vieille tradition dictionnaire comme le français. Au mot *impasse*, par exemple, on trouve en premier lieu, dans le *Petit Larousse* (PL) et le *Petit Robert* (PR) pour ne citer que les plus populaires, le sens « rue, ruelle sans issue » (PL) ou « petite rue qui n'a pas d'issue » (PR), alors que le sens, de loin le plus fréquent, à savoir :

1 Ce mot dénomme une autre réalité en Grande Kabylie : *abandu* (pl. *ibunda*) « Chose située dans la propriété d'un autre et réservée lors d'un partage ou d'une vente : propriété immobilière indépendante du sol (arbre, chemin, parcelle...) ; par extension : chose vendue ou empruntée et qu'on garde » (Dallet : 29). Une autre réalité qui n'a pas d'équivalent simple en français.

« situation ne présentant pas d'issue favorable »² (PL) ou « situation sans issue favorable » (PR), se retrouve en seconde position. Mais il arrive souvent que les deux ordres, logique et fréquentiel, coïncident. Dans l'exemple *azru*, cité ci-dessus, le sens premier du mot, à savoir « caillou ; pierre ; rocher » est aussi le plus fréquent (nous justifions par cette affirmation la remarque de Bernard Al évoquée ici même).

3 Le problème des locutions

L'équivalence est encore plus ardue à établir concernant les locutions, à forte composante culturelle, qui sont spécifiques de chaque langue. On ne peut traduire, par exemple :

[9] *ičča di tbarda-s [iřfa ři řvarđas]* par, littéralement : « il a mangé de/dans son bât »,

énoncé qui a en français un sens exclusivement compositionnel qui n'est pas celui de la locution kabyle, qu'on peut rendre plutôt par « il a trahi les siens ; il agit contre son honneur, contre les siens ».

« La conservation d'une métaphore, au moins en ses éléments principaux, au détriment du sémantisme global implique une inexactitude », écrit Szende (2003 : 265) en citant Lépinette-Sopena. Le sens de cette locution utilisée hors contexte est aujourd'hui difficile à saisir, étant donné le changement du mode de vie dans cette région où l'âne, autrefois utilisé comme bête de somme, a quasiment disparu aujourd'hui, remplacé par les moyens modernes de locomotion. En effet, à défaut de trouver quoi manger et ayant faim, l'âne bête se tourne parfois vers son bât, fabriqué avec des fils d'alfa, pour en manger³. Une locution est une expression linguistique de faits socioculturels propres à une communauté humaine à un moment donné de son histoire. Il existe des dizaines de locutions en kabyle tournant autour de l'âne parce que, jusqu'à voici trente ans, il était un élément important dans la vie quotidienne de la communauté kabyle. La quasi-absence de cet animal aujourd'hui en Kabylie fait que certaines de ces locutions sont de moins en moins utilisées, parce que de plus en plus vagues. Ce flou devient total quand on passe à une autre langue-culture où l'élément en question est totalement inconnu. « Les expressions idiomatiques qui n'ont pas d'équivalent discursif en langue cible devraient être assorties d'une explication plutôt que d'un « équivalent » dont le statut est souvent peu clair pour l'utilisateur », écrit Henri Béjoint (2003 : 218). Par ailleurs, il est préférable dans ces cas, pour l'analyse des locutions en langue source, de donner aussi la traduction littérale qui doit précéder ou suivre celle qui

2 Berbérophone natif, c'est ce deuxième sens que j'ai d'abord connu avant de connaître plus tard le premier qui est rarement utilisé chez nous, en particulier dans la presse écrite et audiovisuelle où le deuxième sens est omniprésent.

3 On raconte que le premier usage de cette locution dans ce sens est fait par le chef d'une tribu de Grande Kabylie en 1857-1858, lors de la conquête française de cette région, en renvoyant un émissaire d'une compagnie militaire française qui lui demandait de les laisser passer pour conquérir une tribu voisine en les faisant bénéficier d'une partie des biens confisqués à celle-ci. Le chef de cette tribu aurait alors renvoyé péremptoirement cet émissaire en le chargeant de dire à ses supérieurs qu'ils n'étaient pas des ânes qui, affamés, mangent (dans) leurs bâts.

convient. Cette traduction permet de reconstituer le sens à partir de ses différents éléments/composants en aidant à comprendre la logique de leur combinaison.

Un autre exemple de locution figée, très utilisée en Petite Kabylie (la région de Béjaïa), est une réponse à la question « comment vas-tu ? » :

[10] *am seksu ddaw teksayt* [æm səksu ddaw θəxsajθ], littéralement : « comme le couscous de dessous la citrouille ».

Une locution dont le sens n'est pas facile à saisir au premier abord même si le couscous est le plat le plus populaire en Kabylie. Il a fallu que j'interroge personnellement certains de ses usagers pour que je comprenne enfin cette réponse qu'on peut traduire par « je ne vais pas bien ; ça ne va pas bien ». C'est que le couscous de dessous la citrouille subit à la fois le poids et la chaleur de cette dernière. Et le fait que la citrouille est aujourd'hui quasiment absente de ce plat, remplacée par la courgette, dont le kabyle a emprunté le nom au français, *takurjiθ* [θakurʒit⁴], complique encore davantage sa compréhension. « Le dictionnaire bilingue décrira le fonctionnement de l'image sans remonter aux sources visuelles, rhétoriques de cette image historiquement et socialement datée et donc en décalage avec la réalité culturelle du moment », écrit Szende (2003 : 266) en citant Robert Galisson.

Par ailleurs, un aspect lexiculturel intéressant de cette locution est qu'elle constitue un euphémisme qui renseigne sur une facette importante du comportement des Kabyles qui font preuve de beaucoup de retenue quand il s'agit de parler d'eux-mêmes, en particulier de leurs difficultés. D'ailleurs, il semble que le mot [11] *sserr* « secret ; discrétion » et par extension « charme ; grâce ; honneur ; bonne réputation, etc. » soit un des mots clés de la langue kabyle qui traduiraient le « tempérament », la « mentalité » ou encore la « sensibilité » de ses locuteurs⁴.

Il recoupe par ailleurs des termes relevant du même type comme [12] *tirugza* [θirugza] (< *argaz* « homme ») « conduite d'un homme qui est digne, respectable et respectueux, juste et loyale envers soi-même et envers les autres, etc. » ; [13] *nnif* « Point d'honneur qui pousse à riposter au défi qui met en cause l'honneur (*lherma*) de la personne ou du groupe : la maison, la femme, les biens, et des biens moraux comme la parole donnée, la foi religieuse, la réputation, etc. ; considération et respectabilité reconnues « à celui qui a assez de point d'honneur pour tenir son honneur à l'abri de l'offense » (Bourdieu⁵) » (Dallet 1982 : 548) ; *lherma* « considération ; respect ; prestige ; bonne renommée » (Dallet 1982 : 337). Tous ces mots, auxquels

4 André Clas et Roda Roberts, se référant à Wandruszka, évoquent ces mots clés concernant quelques langues : « pour l'anglais, ce serait : « humour », « spleen », « common sense », « fair play », « self-control », « understatement » ; pour le français, on aurait : « politesse », « galanterie », « coquetterie », « esprit », « goût », « savoir-vivre », « noblesse », « finesse », « raffinement », « nuance », « verve » ; pour l'italien : « brio », « dolce », « farniente », pour l'allemand : « Gemüt », « Gemütlichkeit », « Stimmung », « Heimweh », « Gründlichkeit », « Erlebnis », « Leistung », pour le portugais « saudade », pour l'espagnol « hidalguia », « grandeza », « machismo », pour le gascon « escarbillat ». » (Clas, A. et Roberts, R. : 241-242).

5 Pour une meilleure compréhension de ce concept de *nnif/nif* on peut se référer aux travaux de Pierre Bourdieu sur la société kabyle, en particulier à *Esquisse d'une théorie de la pratique, précédée de trois études d'ethnologie kabyle* (1972).

on peut ajouter [14] *tissas* « apport; rendement; appui; prestige », [15] *leεqel* « raison; esprit; prudence; sagesse, etc. », [16] *lbaraka* « bénédiction; puissance mystérieuse qui se manifeste par une longue vie, la santé, la fécondité, la prospérité, l'abondance, un profit extraordinaire sur les biens de la terre (*errezq*), les troupeaux, etc. » (Dallet 1982 : 46), constituent de véritables mots clés de la langue-culture kabyle indispensables à son décodage. Deux locutions, choisies parmi tant d'autres, montrent le caractère central du mot *nnif* dans la culture kabyle :

[17] *kulci yettnuz ar nnif* « tout se vend sauf l'honneur » ;

[18] *win ur nesi nnif awer t-id-yaf lexrif* « qui n'a pas d'honneur ne mérite pas d'exister » (litt. qui n'a pas d'honneur qu'il ne vive pas jusqu'à l'automne).

L'importance de cette valeur est exprimée même, par euphémisme, dans la première locution où celui « qui agit contre son honneur » est comparé à l'âne, animal le plus déconsidéré dans la littérature kabyle où il symbolise la bêtise et la stupidité. D'ailleurs on ne trouve pas dans cette langue de qualificatif plus méprisable à son endroit que son nom lui-même : [19] *ayyul d ayyul!* « l'âne est âne ! ».

Il paraît donc évident à travers ces exemples que la locution, comme l'écrit si bien Catherine Caws (2006 : 206), est « très souvent témoin d'une croyance ou d'une culture dont elle affiche les propriétés symboliques » qu'il convient de prendre en charge dans l'établissement de l'équivalence, même si la tâche semble a priori malaisée.

4 Le problème de l'équivalence d'usage

Une bonne équivalence ne doit pas être réduite au strict plan sémantique, c'est-à-dire au niveau de la langue. Elle doit traduire l'usage dans sa dimension sociolinguistique et culturelle. Les marques d'usage, utilisées par les lexicographes, ont pour fonction de préciser le type d'usage qui est fait d'un mot ou d'une expression et permettent aux usagers des dictionnaires de passer de la langue au discours, de l'abstrait au concret. Leur prise en charge dans l'équivalence, en particulier en lexicographie active, est nécessaire. Ces marques ne doivent pas cependant révéler « la position du lexicographe-locuteur par rapport à la langue et au monde » (Glatigny 1990 : 7), mais elles doivent plutôt révéler celle des unités concernées par rapport à la « langue neutre » ou « passe-partout ». Mais le caractère flou et subjectif de la notion d'usage neutre, par rapport auquel elles sont déterminées et qui peut être interprété différemment par les lexicographes, font qu'elles ne correspondent pas toujours à l'expérience propre des locuteurs et qu'elles sont parfois employées différemment, pour une même unité, d'un dictionnaire à l'autre et d'une édition à l'autre pour un même dictionnaire (v. Girardin 1987 : 78). S'ajoute à cela le problème de la non-équivalence des normes. La notion d'usage neutre ou de norme concerne une langue et une communauté linguistique bien déterminées. Dans la tradition lexicographique française, c'est l'usage de la bourgeoisie ou celui de la cour qui étaient neutres, aujourd'hui celui des écrivains et des journalistes, alors

que celui du peuple, de la masse, est marqué (*populaire*). Cette norme ne peut pas être celle du berbère, langue vernaculaire et à tradition orale, donc foncièrement populaire, seul usage qu'elle a.

Pour revenir aux locutions, il serait particulièrement intéressant, du moins dans les dictionnaires de thème, de les traduire par d'autres locutions, dans la mesure de leur disponibilité. Un usager actif qui voudrait produire ou traduire dans la langue cible, notamment en littérature, préférerait certainement des locutions, même avec un sémantisme approximatif, aux simples mots. L'essentiel est que le sémantisme global de la locution soit conservé, sinon on serait dans l'« inexactitude », pour reprendre le terme de Lépinette-Sopena (Szende *ibid.*). Il ne s'agit cependant pas de faire de « belles infidèles » à la Perrot d'Ablancourt « où l'agrément du style est privilégié au mépris de la fidélité au texte d'origine » (Pruvošt 2003 : 2). La meilleure équation d'équivalence à notre avis est celle qui met en rapport l'ensemble des aspects d'un élément discursif : le sémantique, le stylistique et le pragmatique.

Une locution proverbiale kabyle comme [20] *ulac ticaḍ mebla idammen*, litt. « il n'est pas de tatouage sans sang », peut être rendue par une autre locution française équivalente au triple plan sémantique, stylistique et pragmatique, à savoir qu'« on ne fait pas d'omelette sans casser des œufs ». Le proverbe [21] *ur ttamen ššaba ar't-terwet* « litt. ne crois à la bonne récolte qu'après le battage », peut aussi être rendue par un autre proverbe bien connu en français, en l'occurrence : « il ne faut pas vendre la peau de l'ours avant de l'avoir tué » ; [22] *win yeqqes wezrem yettagad aseḡwen* « litt. celui qu'a piqué le serpent craint une corde (d'alfa) » est équivalent au proverbe français : « chat échaudé craint l'eau froide » ; etc. C'est cette équivalence qui ménage sens, style et condition d'usage, qui n'est cependant pas facile d'accès, qu'on peut appeler, avec Antoine Berman, cité par Vrinat-Nikolov, la « mise en rapport des langues et des cultures » : « L'essence de la traduction est d'être ouverture, métissage, décentrement : Elle est mise en rapport ou elle n'est rien » (Vrinat-Nikolov 2003 : 135).

5 Le problème du classement des locutions en macrostructure

Se pose aussi le problème du classement des locutions en nomenclature et celui de leur hiérarchie en microstructure. Si cette dernière peut suivre la même démarche que celle établie pour les différentes acceptions d'un même mot (voir point 2), le problème de leur rattachement à l'un de leurs constituants reste entièrement posé. Dans le dictionnaire kabyle-français de J.-M. Dallet, faute d'une méthode de classement de ces locutions, on les retrouve parfois sous plusieurs entrées, au détriment de l'aspect dictionnaire⁶ de l'entreprise qui recommande l'économie de l'espace pour un meilleur coût de production de l'ouvrage. On retrouve ainsi la locution [23]

6 « Lorsque la notion de lexicographie cesse d'impliquer la réalisation d'un dictionnaire, il faut bien dénommer la partie dissociée du concept unitaire antérieur. Or cette partie constitue un domaine complexe et bien circonscrit, celui du *Dictionnaire*, représentant l'ensemble de tous les types de dictionnaires et de tout ce qui s'y rapporte. Ce n'est pas rien. Pour désigner le champ couvert à la fois par le *dictionnaire-réalisation* et le *dictionnaire-objet d'étude*, j'avais choisi *dictionnaire* » (Quémada 1987 : 235).

amcic yennumen tuggi leqrar-is/d ulaqrar ad yettwiṭṭef « le voleur finit toujours par se faire prendre » ou, littéralement, « le chat habitué à la marmite finira par se faire prendre » sous les entrées pratiquement de tous les constituants de la locution : sous *leqrar* (p. 671), *tuggi* (p. 247), *ttwiṭṭef* (p. 836) et sous *nnam* (p. 565) ! La locution [24] *ččiy iḥder uwackan, d iyil-iw i yi-t-yefkan* « j'ai gagné mon argent, j'ai bien le droit de le dépenser » ou, littéralement, « j'ai mangé un morceau de galette, c'est mon bras qui me l'a procuré » se retrouve sous ses trois premiers constituants : sous *ečč* (p. 69), *uwackan* (p. 850) et sous *iḥder* (p. 308) ; [25] *lemleḥ ufus* « pourboire, gratification » ou, littéralement, « le sel de la main » sous ses deux constituants ; etc. Pour Robert Martin « la locution gagne à être traitée sous le vocable qui s'éloigne le plus de son sens habituel ; ainsi *baptiser le vin* « le couper d'eau » sera mentionné de préférence sous *baptiser* » (Martin 2000 : 57). Le sens le moins habituel d'un mot est effectivement celui qui a le plus besoin d'être illustré. Mais il n'est pas toujours évident de savoir lequel des vocables est celui qui s'éloigne le plus de son sens habituel. Dans le proverbe :

[26] *Lamana fell-as ay terya tgerfa* (litt. « le dépôt confié c'est à cause de lui qu'a brûlé le corbeau ») « c'est à cause de la confiance (trahie) que le corbeau a brûlé »,

lequel de *lamana* « dépôt confié » qui devient « confiance » et de *terya* < *ery* « brûler » qui devient « noircir » s'éloigne plus de son sens habituel ?

On peut combiner cette suggestion de R. Martin avec celle qui voudrait qu'on considère le sens le mieux illustré ou qui constitue le thème (central) de la locution. Ainsi, la locution ci-dessus sera traitée sous *lamana* qui en constitue le sujet. On peut compléter ce choix par l'ajout du traitement de la locution, dans la mesure de la disponibilité de l'espace dans la version papier, sous les mots (ou les sens) d'occurrence rare qui ont très peu ou pas du tout d'autres contextes illustratifs. Dans la deuxième locution citée dans cette partie, le mot (*iḥder*) *uwackan*⁷ « (morceau) de galette » qui est d'occurrence rare peut recevoir le traitement de cette locution en plus du mot *ečč* « manger » qui en constitue le thème.

On peut convenir aussi de traiter une locution plus simplement sous son premier constituant, ou en choisissant un ordre préférentiel de ceux-là en ne tenant pas compte des mots grammaticaux, comme le fait le DMF (*Dictionnaire du moyen français*), « sous le premier substantif qu'elles comportent (de la gauche vers la droite), à défaut sous le premier adjectif, à défaut le premier adverbe, à défaut le premier verbe » (Martin 2000 : 57). Méthode très pratique, mais comportant l'inconvénient de « surillustrer » certains mots en en « sous-illustrant » d'autres. On peut aussi combiner diverses solutions selon les cas, mais la démarche choisie doit s'appliquer de façon cohérente à l'ensemble des locutions traitées. Cela fera gagner beaucoup d'espace et faire baisser le prix coûtant de l'ouvrage, en particulier pour les versions papier. Pour la version électronique, l'espace est par contre disponible, mais on risque d'encombrer presque inutilement l'article en multipliant les locutions et de rendre ainsi leur consultation plus difficile.

7 C'est son synonyme *ayrum* (*aquran/uṭajin*) « pain (dur/du poêlon de terre) » qui est utilisé dans le langage courant.

Conclusion

Un dictionnaire bilingue de langue doit être un outil d'aide à l'intercompréhension et à la communication interculturelle. Pour être efficace, il doit donner l'information qu'il faut à la place qu'il faut. Une information de trop est une information qui coûte en temps et même en argent à l'utilisateur. Elle n'est pas inutile, elle est nuisible. Une information en moins peut être source de confusion et d'incompréhension et rendrait l'usage même du dictionnaire problématique. Aussi l'utilisateur doit-il pouvoir trouver l'information qu'il recherche facilement en suivant un schéma cohérent de l'organisation de celle-ci à l'intérieur du dictionnaire. Une disposition qui doit être nécessairement contraire à celle des livres d'une bibliothèque où « les moins utiles sont les plus haut placés ».

Bibliographie

- Béjoint (H.), 2003 : « Vers un dictionnaire bilingue de médiation », dans Szende (Th.), dir., *Les écarts culturels dans les dictionnaires bilingues*, Paris, Honoré Champion, p. 207-222.
- Bourdieu (P.), 1972 : *Esquisse d'une théorie de la pratique, précédée de trois études d'ethnologie kabyle*, Genève-Paris, Droz.
- Caws (C.), 2006 : « Nouvelle approche du traitement lexicographique de la locution : l'exemple de entre chien et loup », dans *Cahiers de lexicologie*, n° 89, Éditions Garnier, p. 195-207.
- Clas (A.) et Roberts (R.), 2003 : « Le dictionnaire bilingue : une mosaïque culturelle ? », dans Szende (Th.), dir., *Les écarts culturels dans les dictionnaires bilingues*, Paris, Honoré Champion, p. 237-248.
- Dallet (J.-M.), 1982 : *Dictionnaire kabyle-français (parler des Ait Menguellat, Algérie)*, Paris, SELAF, 1 052 p. + XXXIX.
- Duval (A.), 2003 : « Nul ne peut-il ignorer la loi dans les équivalences culturelles entre le français et l'anglais ? », dans Szende (Th.), dir., *Les écarts culturels dans les dictionnaires bilingues*, Paris, Honoré Champion, p. 65-74.
- Galisson (R.), 1987 : « Accéder à la culture partagée par l'entremise des mots à CCP », dans Galisson (R.) et Lerat (P.), coord., *Le poids des mots dans l'enseignement/apprentissage du F.L.E.*, éla n° 67, Paris, Didier Erudition, p. 119-140.
- Girardin (Ch.), 1987 : « Système des marques et connotations sociales dans quelques dictionnaires culturels français », dans *Lexicographica*, n° 3, p. 76-102.
- Glatigny (M.), 1990 : « Présentation : l'importance des marques d'usage », dans *Lexique*, n° 9, p. 7-16.
- Huyghe (P. G.), 1904 : *Dictionnaire kabyle-français*, 2^e édition, Paris, Imprimerie Nationale.
- Landheer (R.), 1981 : « Ambiguïté et dictionnaire bilingue », dans *Le dictionnaire, Actes du colloque franco-néerlandais* 28-29 avril 1981, Presses Universitaires de Lille, p. 147-156.
- Le Petit Larousse Compact*, 2002, Paris, Larousse.
- Le Petit Robert (format électronique)*, 2002, Dictionnaire analogique et alphabétique de la langue française, texte remanié et amplifié sous la dir. de Rey-Debove (J.) et Rey (A.), Paris, nouvelle édition Petit Robert de Paul Robert.

- Martin (R.), 2000 : « Perspectives en lexicographie informatisée : l'expérience du DMF (Dictionnaire du Moyen Français) », dans *Mémoires de la société de linguistique de Paris*, nouvelle série T. VII, Lexique, lexicologie, lexicographie, Peeters, Louvain (Belgique), p. 51-71.
- Pruvošt (J.), 2003 : « Avant-Propos. De l'« infidèle » traduction à la mention des écarts culturels dans les dictionnaires bilingues », dans Szende (Th.), dir., *Les écarts culturels dans les dictionnaires bilingues*, Paris, Honoré Champion, p. 1-4.
- Pruvošt (J.), 2005 : « Quelques concepts lexicographiques opératoires à promouvoir au seuil du XXI^e siècle », dans *Revue de dialectologie des langues-cultures*, « Dictionnaires et innovation », éla n° 137 janv.-mars 2005, p. 7-37.
- Quémada (B.), 1987 : « Notes sur lexicographie et dictionnaire », dans *Cahiers de lexicologie*, n° 51, Paris, Didier Erudition, p. 235-245.
- Rahmani (S.), 1934 : *Coutumes kabyles du Cap-Aokas : La grossesse, la naissance et la vie de l'enfant jusqu'à la circoncision*, Alger, Société Historique Algérienne.
- Song (E.-J.), 2003 : « Lexiculture et pragmaculture dans un dictionnaire bilingue d'énoncés usuels (coréen, français) », dans Szende (Th.), dir., *Les écarts culturels dans les dictionnaires bilingues*, Paris, Honoré Champion, p. 365-382.
- Szende (Th.), 2000 : « L'information sémantique en lexicographie bilingue (hongrois-français) », dans Szende (Th.), dir., *Dictionnaires bilingues : méthodes et contenus*, Paris, Honoré Champion, p. 69-81.
- Szende (Th.), 2003 : « Images collectives et formules codées », dans Szende (Th.), dir., *Les écarts culturels dans les dictionnaires bilingues*, Paris, Honoré Champion, p. 259-270.
- Vrinat-Nikolov (M.), 2003 : « Heurs et malheurs des traducteurs face aux dictionnaires bilingues », dans Szende (Th.), dir., *Les écarts culturels dans les dictionnaires bilingues*, Paris, Honoré Champion, p. 135-146.

*Heidi Verplaetse,
Cornelia Wermuth,
Philippe Anckaert,
June Eyckmans,
Winibert Segers,*
Lessius/KULeuven/HEFF Bruxelles/Erasmushogeschool Brussel/VUB,
Belgique,
winibert.segers@lessius.eu.

La qualité de la traduction médicale et les exercices terminologiques préparatoires

Résumé

À l'occasion d'une expérience, nous avons étudié l'influence d'un exercice terminologique préparatoire sur la qualité de traductions faites par des étudiants. Le groupe expérimental a traduit un texte médical après avoir fait un exercice terminologique préparatoire. Le groupe de contrôle a traduit le même texte sans préparation terminologique. En évaluant les traductions selon une méthode purement formelle, nous n'avons pu déceler aucune influence positive de la préparation terminologique sur la qualité de ces traductions.

Mots-clés: didactique; évaluation; méthode formelle; traduction médicale.

La question que nous nous sommes posée est simple: quelle est l'influence d'un exercice terminologique préparatoire sur la qualité d'une traduction? Cette question est issue de notre pratique didactique en matière de traduction médicale. Dans nos cours, nous recourons en effet à des exercices terminologiques préliminaires à l'exercice de traduction proprement dit. Ces exercices terminologiques consistent le plus souvent en une analyse morphologique de termes médicaux, axée sur les racines et les affixes, suivie de leur traduction. L'idée sous-jacente est la suivante: un exercice terminologique préparatoire a une influence positive sur la qualité d'une traduction parce qu'il permet aux étudiants de mieux comprendre les termes et le contenu du texte à traduire. Nous avons essayé de vérifier cette hypothèse dans le cadre d'une expérience, en demandant à deux groupes d'étudiants de traduire un texte médical: le groupe expérimental, constitué de 12 étudiants, après avoir fait un exercice terminologique préparatoire, le groupe de contrôle, constitué de 13 étudiants, sans avoir bénéficié d'une préparation terminologique. Tous les sujets d'expérience ont traduit le même texte, du français en néerlandais, c'est-à-dire dans leur langue maternelle. Il s'agissait d'étudiants en quatrième et dernière année de la formation au métier de traducteur. Pour cette expérience, nous avons fait appel à tous les étudiants inscrits au cours de traduction médicale français-néerlandais. Ils ont pu utiliser toutes les ressources d'information (dictionnaires, glossaires,

Internet, etc.) qu'ils jugeaient nécessaires. Le texte qu'ont traduit les deux groupes s'intitulait « Hypoglycémie: qu'est-ce que c'est ? ». Il s'agissait d'un texte destiné à un public non spécialisé, que nous avons trouvé sur le site *PasseportSanté.net* et que nous reproduisons ci-dessous :

Hypoglycémie: qu'est-ce que c'est ?

On dit qu'une personne souffre d'hypoglycémie quand elle présente un ensemble de symptômes qui sont le résultat d'une baisse de la glycémie, ou taux de glucose dans le sang (appelé communément « taux de sucre »). Plus précisément, la glycémie doit être inférieure à 3,5 millimoles par litre de sang (mmol/l) pour que le diagnostic d'une véritable hypoglycémie soit posé.

L'hypoglycémie est parfois difficile à cerner. Ses symptômes, en plus d'être peu spécifiques, peuvent varier beaucoup d'une personne à l'autre. Les plus fréquents sont la baisse d'énergie et la nervosité. Cependant, certaines personnes n'ont aucune fatigue, mais une vision trouble ou une difficulté à se concentrer. Les symptômes de l'hypoglycémie réactionnelle apparaissent le plus souvent de trois à quatre heures après un repas.

Il faut faire attention de ne pas baptiser « hypoglycémie » un état qui pourrait être la conséquence d'une fatigue, d'un stress ou d'une nervosité qui déstabilise¹.

Le texte retenu pour l'expérience est plutôt bref et simple sur le plan conceptuel, ne posant normalement pas de problème de compréhension à des lecteurs non spécialisés. Il correspond au type de textes que les étudiants sont amenés à traduire au cours de traduction médicale, du français vers le néerlandais, à savoir des textes destinés à un large public sans connaissances médicales particulières. Ce texte contient les termes médicaux suivants: « hypoglycémie », « glycémie », « taux de glucose », « hypoglycémie réactionnelle ».

L'exercice terminologique préparatoire, fait par le groupe expérimental, consistait à répondre à six questions :

1. Quels termes médicaux trouvez-vous dans le texte ?
2. Quels sont les racines et affixes de ces termes médicaux ?
3. Que signifient ces racines et affixes ?
4. Connaissez-vous d'autres termes avec la même racine ?
5. Connaissez-vous d'autres termes avec les mêmes affixes ?
6. Quelle est la traduction néerlandaise des termes médicaux français que vous avez trouvés dans le texte ?

1 www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=hypoglycemie_pm#P25_1231

Plutôt que d’embrasser tous les aspects de la terminologie médicale, nous avons délibérément limité les exercices de préparation terminologique à certains aspects morphologiques afin de limiter le nombre de variables impliquées dans l’expérience.

Cet exercice préparatoire a-t-il une influence sur la qualité des traductions ? Afin de répondre à cette question, il fallait opter pour une méthode d’évaluation de la qualité des traductions fournies par les étudiants. Dans la littérature, on distingue principalement deux types d’évaluation : la méthode dite « holistique » et la méthode analytique (Waddington 2001 ; Anckaert, Eyckmans et Segers 2008). Dans le cas de la méthode holistique, l’enseignant-évaluateur lit chacune des traductions dans son intégralité. Il attribue ensuite une note reflétant son impression globale de la qualité, sans détailler explicitement pourquoi ni comment il arrive à cette note. À l’opposé, la méthode analytique implique le recours à une grille d’évaluation préétablie, comprenant un nombre de critères réputés rendre compte de l’éventail des différentes erreurs de traduction possibles : grammaticales, lexicales, d’orthographe, de style, etc. Chaque erreur est pénalisée par un retrait de points fixé *a priori* (-1, -2, -3, etc.) et à chaque catégorie d’erreurs est affecté un facteur de pondération également préétabli (grammaire : 4/20 ; lexique : 4/20 ; orthographe 4/20 ; style 4/20 ; etc.). Ces deux méthodes d’évaluation étant tributaires de la subjectivité de l’individu qui évalue, notamment de sa conception toute personnelle de ce qu’est soit une bonne traduction, soit une erreur de tel ou tel type, mais également du caractère éminemment faillible de l’évaluateur (risques d’incohérence, effets de halo ou de séquence), elles ne vont pas sans poser des problèmes tant dans les milieux professionnels et scolaires (sélection de candidats, examens, etc.) que dans le contexte d’une recherche scientifique. Afin d’éviter les problèmes de subjectivité et d’infidélité des méthodes holistiques et analytiques, nous avons choisi d’utiliser une méthode d’évaluation purement formelle, basée sur l’outil d’assurance qualité *QA Distiller* développé par Yamagata Europe (www.qa-distiller.com/). Les concepteurs de cet outil fondent leur approche sur l’hypothèse (non démontrée) selon laquelle les erreurs formelles fourniraient une indication fiable de la qualité globale de la traduction.

Nous donnons deux exemples d’erreurs formelles :

[1] Fr. Hypoglycémie > Nl. *Hyoglycemie*

[2] Fr. 3,5 millimoles > Nl. 35 *millimol*

Il faut préciser qu’il s’agit dans notre recherche de la qualité globale du groupe, de la performance du groupe et non pas de la performance individuelle.

Voici les résultats du groupe expérimental :

7 erreurs formelles

[3] *sysmptomen* (*symptomen*) (étudiant 3)

[4] *hypoglycaemie* (*Hypoglaecemie*) (étudiant 3)

[5] *zenuwachtig heid* (*zenuwachtigheid*) (étudiant 3)

[6] *Hypoglykemie (hypoglykemie)* (étudiant 4)

[7] *Hyoglycemie (Hypoglycemie)* (étudiant 12)

[8] *mmol/(mmol/l)* (étudiant 12)

[9] *hypoglycemie (hypoglycemie)* (étudiant 12)

Et les résultats du groupe de contrôle :

7 erreurs formelles

[10] *hypoclaecemie (hypoglaecemie)* (étudiant 1, deux fois)

[11] *lage (laag)* (étudiant 8)

[12] *Hypoglycemie (hypoglycemie)* (étudiant 10, quatre fois)

Il est remarquable que le plus grand nombre d'erreurs porte sur le terme clé du texte: la traduction du terme « hypoglycémie ».

Voici les détails pour le groupe expérimental :

Nombre de fautes formelles

| | |
|-------------|---|
| Étudiant 1 | 0 |
| Étudiant 2 | 0 |
| Étudiant 3 | 3 |
| Étudiant 4 | 1 |
| Étudiant 5 | 0 |
| Étudiant 6 | 0 |
| Étudiant 7 | 0 |
| Étudiant 8 | 0 |
| Étudiant 9 | 0 |
| Étudiant 10 | 0 |
| Étudiant 11 | 0 |
| Étudiant 12 | 3 |

Et les détails pour le groupe de contrôle :

Nombre de fautes formelles

| | |
|-------------|---|
| Étudiant 1 | 2 |
| Étudiant 2 | 0 |
| Étudiant 3 | 0 |
| Étudiant 4 | 0 |
| Étudiant 5 | 0 |
| Étudiant 6 | 0 |
| Étudiant 7 | 0 |
| Étudiant 8 | 1 |
| Étudiant 9 | 0 |
| Étudiant 10 | 4 |
| Étudiant 11 | 0 |
| Étudiant 12 | 0 |
| Étudiant 13 | 0 |

| | | |
|---------------------|-------|-----------------------|
| – Étudiant 5 | 36/65 | |
| – Étudiant 6 | 34/65 | |
| – Étudiant 7 | 45/65 | |
| – Étudiant 8 | 46/65 | |
| – Étudiant 9 | 39/65 | |
| – Étudiant 10 | 41/65 | |
| – Étudiant 11 | 41/65 | |
| – Étudiant 12 | 36/65 | (3 erreurs formelles) |

Si nous comparons les résultats de l'évaluation formelle à ceux des évaluations holistiques, nous observons que :

- l'étudiant 2 a la note la plus basse (30/65), mais cet étudiant n'a fait aucune faute formelle ;
- l'étudiant 3 occupe la onzième, l'avant-dernière place (31/65) et a fait trois fautes formelles ;
- l'étudiant 12 occupe la huitième/neuvième place (36/65) et a fait trois fautes formelles ;
- l'étudiant 4 occupe la deuxième/troisième place (45/65) et a quand même fait une faute formelle.

Alors que les positions des étudiants 3 et 12 semblent confirmer l'idée sous-jacente de l'évaluation formelle selon laquelle des erreurs formelles sont indicatives de la qualité globale de la traduction, les positions des étudiants 2 et 4 viennent ébranler cette hypothèse.

Par ailleurs, les évaluations holistiques illustrent à merveille la diversité des attitudes que peuvent adopter des individus lors d'une même tâche, à savoir l'évaluation de la qualité globale d'une traduction. Alors que l'étudiant-évaluateur H, par exemple, n'attribue jamais que deux notes (2/5 et 3/5), l'étudiant-évaluateur A parcourt presque toute l'échelle de valeurs, en discernant des notes allant de 1 à 5. Les jugements de valeur émis lors d'une évaluation holistique peuvent en outre donner lieu à des divergences remarquables : la traduction de l'étudiant 5 est évaluée à un modeste 1/5 par l'étudiant C, alors qu'elle décroche la note maximale (5/5) lorsqu'elle est jugée par l'étudiant-évaluateur I.

La terminologie fait partie intégrante de la formation des traducteurs et donne lieu à diverses batteries d'exercices, mais à la lumière des résultats obtenus, nous posons la question de leur utilité, du moins dans la forme que revêt actuellement ce type de préparation terminologique. Notre conclusion sur l'absence d'influence des exercices terminologiques préparatoires sur la qualité globale d'une traduction doit être interprétée avec prudence, d'une part parce qu'elle a trait à une préparation limitée aux aspects morphologiques de la terminologie médicale et d'autre part parce que les résultats de notre expérience ont été obtenus avec un échantillon de population pour le moins modeste, en appliquant une méthode purement formelle. Cette dernière garantit certes une évaluation objective, mais ne prend

peut-être pas en compte des aspects importants de la qualité globale d'une traduction. Notre étude montre une fois de plus (PACTE 2000; Eyckmans, Anckaert et Segers 2009) que la mise au point d'une méthode d'évaluation fidèle et valide devrait constituer une priorité en traductologie. Sans instruments de mesure adéquats, c'est-à-dire répondant aux deux critères docimologiques de base (fidélité et validité), la recherche empirique dans notre discipline est condamnée au piétinement, au raisonnement circulaire, voire à l'inaction.

Bibliographie

- Anckaert, (Ph.), Eyckmans, (J.) et Segers, (W.), 2008 : « Pour une évaluation normative de la compétence de traduction », dans *ITL Review of Applied Linguistics*, vol. 41, n° 155. p. 53-76.
- Andriesen, (S.), 2008 : « Quality Control in Medical Translations », www.medilingua.com/.../QualityControlInMedicalTranslation_IPT2008.pdf
- Applegate, (A.) et Overton, (V.), 1994 : *The Elements of Medical Terminology*, Albany (N.Y.), Delmar.
- Balliu, (Ch.), 1994 : « L'enseignement de la traduction médicale: pour une nouvelle pragmatique », dans *Meta*, vol. 39, n° 1. p. 15-25. www.erudit.org/revue/meta/1994/v39/n1/001964ar.pdf
- Balliu, (Ch.), 2005 : « La didactique de la traduction médicale, deux ou trois choses que je sais d'elle », dans *Meta*, vol. 50, n° 1. p. 67-77. www.erudit.org/revue/meta/2005/v50/n1/010658ar.pdf
- Eyckmans, (J.), Anckaert, (Ph.), Segers, (W.), 2009 : « The perks of norm-referenced translation evaluation », dans Angelelli, (Cl.), Jacobson, (H.), éd., *Testing and Assessment in Translation and Interpreting Studies. A call for dialogue between research and practice*, Amsterdam, John Benjamins, p. 73-93.
- Hoof, (H.) van, 1986 : *Précis pratique de traduction médicale (anglais-français)*, Paris, Maloine.
- Hutton, (A.R.), 1999 : *An Introduction to Medical Terminology: a Self-teaching Package*, Edinburgh, Churchill Livingstone.
- Lankamp, (R.E.), 1988 : *A Study on the Effect of Terminology on L2 Reading Comprehension. Should Specialist Terms in Medical Texts be Avoided?*, Amsterdam, Rodopi.
- Lee-Jahnke, (H.), 2001 : « L'enseignement de la traduction médicale: un double défi ? », dans *Meta*, vol. 46, n° 1, p. 145-153. www.erudit.org/revue/meta/2001/v46/n1/003445ar.pdf
- PACTE, 2000 : « Acquiring Translation Competence. Hypotheses and Methodological Problems of a Research Project », dans Beeby, (A.), Ensinger, (D.) et Presas, (M.), éd., *Investigating Translation*, Amsterdam/Philadelphia, John Benjamins, p. 99-106. www.fti.uab.es/pacte/publicacions/Benjamins2000.pdf
- Rouleau, (M.), 1994 : *La traduction médicale. Une approche méthodique*, Brossard, Linguattech.
- Rouleau, (M.), 1996 : *La traduction médicale. Livre du maître*, Brossard, Linguattech.
- Waddington, (Ch.), 2001 : « Different Methods of Evaluating Student Translations: The Question of Validity », dans *Meta*, vol. 46, n° 2. p. 311-325. www.erudit.org/revue/meta/2001/v46/n2/004583ar.pdf

Madalena Contente,
Centro de Linguística da Universidade Nova de Lisboa,
Departamento de Educação Médica,
Faculdade de Ciências Médicas – Universidade Nova de Lisboa,
Lisboa,
Portugal,
madalena.contente@fcm.unl.pt.

Sebastião Silva Filho,
Centro de Linguística da Universidade Nova de Lisboa,
Lisboa,
Portugal,
sebafilho@hotmail.com.

Ontoterminologie de l'éducation médicale : conception et organisation

Résumé

L'ontoterminologie de l'éducation médicale, représentation de ce domaine de connaissance, a comme but final l'aide à la formation des médecins-enseignants des Facultés de médecine. Ce projet contribuera à l'autonomie et à l'auto-apprentissage de ces professionnels ainsi qu'à l'intercommunication scientifique. Les résultats seront intégrés dans des modules de formation sur les différentes tâches pédagogiques des médecins-enseignants qui veulent développer une autoformation. L'autonomie scientifique et pédagogique de l'enseignement supérieur en médecine exige des médecins-enseignants une double compétence : la maîtrise scientifique de leur spécialité ainsi que des compétences sur les méthodologies d'enseignement de leurs disciplines.

Mots-clés : ontologie ; terminologie ; ontoterminologie ; éducation médicale.

Introduction

L'ontoterminologie de l'éducation médicale, objet de notre recherche, élaborée à partir de corpus, a comme but final l'aide à la formation des médecins-enseignants des Facultés de médecine et des enseignants d'autres institutions supérieures de sciences de la santé. La modélisation du domaine est présentée à travers une ontologie articulée avec une terminologie relative aux concepts de l'éducation médicale qui sont utilisés par une communauté scientifique. L'extraction de termes ainsi que l'identification des différents types de relations conceptuelles sont effectuées à l'aide du traitement semi-automatique du corpus associé aux méthodologies de

la terminologie computationnelle. L'outil informatique, élaboré spécifiquement pour ce travail, sera intégré dans des modules d'autoformation, visant l'appropriation et la compréhension des concepts du domaine de l'éducation médicale. Cette recherche s'appuie sur un corpus de textes spécialisés d'éducation médicale. Nous avons pu compter sur la collaboration d'un ingénieur de la connaissance (Rui Gorgueira du *Centro de Linguística da Universidade Nova de Lisboa*) et d'un spécialiste du domaine (Professeur Patrícia Rosado Pinto de la *Faculdade de Ciências Médicas de l'Universidade Nova de Lisboa*).

1 Ontologie : les différentes approches

La recherche sur les ontologies s'est développée en fonction d'un travail interdisciplinaire en sciences cognitives, en logique, en intelligence artificielle, en ingénierie des connaissances, en informatique et en terminologie : Guarino (1997, 1998), Gruber (1995), Sowa (1994, 2000, 2006), Fürst (2004). Selon le document ISO 24 156 *Guidelines for Applying Concept Modeling in Terminology Work* (cf. ISO 704 et ISO 1087), ontologie et terminologie sont interassociées à travers le concept de modélisation du domaine. Le concept d'ontologie, issu de la philosophie, a été repris par l'ingénierie des connaissances avec une nouvelle acception désignant la représentation formelle des systèmes notionnels et, d'une façon particulière, par la terminologie tant comme un point de vue méthodologique, que comme une application au traitement de l'information.

Guarino (1995) définit l'ontologie en tant que domaine de la partie de la philosophie qui s'intéresse à la nature et à l'organisation de la réalité. Selon Sowa (2000), l'ontologie en tant que domaine est l'étude des catégories d'entités abstraites et concrètes qui existent ou peuvent exister ; c'est à partir de ces catégories que l'on peut construire les graphes conceptuels ; dans cette perspective l'ontologie est une abstraction qui contient la représentation des connaissances d'un domaine. Ontologie et terminologie présupposent la conceptualisation d'un domaine du savoir ; elles visent les mêmes propriétés de consensus, de cohérence, de partage et de réutilisation. L'ontologie, en ingénierie des connaissances, est définie comme une conceptualisation d'un domaine partagée par une communauté scientifique et/ou professionnelle, c'est-à-dire comme un ensemble de concepts et de relations interconceptuelles définies à l'aide d'un langage formel compréhensible par un ordinateur (Roche 2007). L'ontologie est une représentation sous différentes perspectives des propriétés de l'objet, conceptualisé à travers des concepts dénommés par les langues.

Ces dernières années, l'approche ontologique s'est enrichie grâce à de nouveaux outils et de nouveaux langages informatiques ; l'ontologie emprunte à la terminologie l'objet d'étude, l'approche conceptuelle, les relations entre concepts, les relations entre termes et concepts, les propriétés des concepts, l'analyse de la désignation et les méthodes de classification, en particulier, celles des domaines. Elle emprunte à la terminologie la structuration des termes, l'analyse de la signification

du terme, l'analyse du discours scientifique (la segmentation des unités du discours, les liaisons entre les segments du discours). Il n'existe pas de consensus sur le terme ontologie, car tous les systèmes de représentation du monde sont symboliques avec une dimension épistémique. Il est difficile d'attribuer aux êtres un statut ontique, indépendant de la vision du monde, de la culture ou de la théorie scientifique qui les rend intelligibles. Décrire les êtres présuppose un point de vue sur les langues naturelles. Par la façon dont ils représentent la réalité, les différents lexiques ne configurent pas une vision ontologique et universelle comme en attestent des études de linguistique comparée, de traductologie et de traduction automatique.

Les ontologies épistémiques (Barry Smith 1998 : 109) constituent des points de vue (conceptualisations) selon lesquels un individu, un groupe, un langage ou une science conceptualisent le monde ou un domaine particulier (Bento 2007). Une ontologie implique ou comprend un certain point de vue du monde par rapport à un domaine de la connaissance; ce point de vue est souvent conçu comme un ensemble de concepts – p. ex., entités, attributs, processus –, leurs définitions et leurs interrelations. On appelle cela une conceptualisation.

Une ontologie peut prendre différentes formes, mais elle inclut nécessairement un vocabulaire de termes et une spécification de leur signification. Selon Charlet (2007), une ontologie est une spécification qui rend partiellement compte d'une conceptualisation.

D'après Depecker (2007 : 17), « on dessine aujourd'hui deux grandes tendances : l'ontologie à base textuelle; l'ontologie à base conceptuelle, de type bases de connaissances développées ». Cet auteur considère que, pour l'ontologie, il y a deux types d'approches : la construction d'ontologies à partir de textes et la construction d'ontologies à partir de concepts.

Un modèle conceptuel avec une composante terminologique construite à partir des données lexicales et linguistiques est consigné dans une fiche qui nous permet de passer d'un niveau terminologique à un niveau conceptuel. Une première phase consiste dans l'identification des termes centraux du domaine et dans leur organisation dans des hiérarchies, tout en faisant le repérage des concepts de façon à formuler la modélisation du domaine. Une fois les concepts délimités et stabilisés, on les définit et on les intègre dans des hiérarchies ou des structures; puis on les regroupe de façon à créer de nouveaux ensembles partageant des propriétés particulières (Contente 2008).

Dans la bibliographie internationale, on peut trouver plusieurs types d'ontologies médicales ayant des objectifs et des fonctions différents qui répondent à des besoins scientifiques, documentaires et/ou sociaux: Charlet, J. (2007), Zweigenbaum, P., Charlet, J. (2007), Zweigenbaum, P., Courtois, P. (1998).

Cependant, il y a d'autres ontologies dans le domaine de l'éducation médicale qui s'occupent de la gestion autonome du plan d'études par l'étudiant; d'autres ont

comme objet des spécialités médicales: Bratsas, C. Kaimakamis, E., Koutkias, V. Bamidis, P.D., Maglaveras, N. (2007); Ukor, R. et Armitt, G. (2008).

2 Ontoterminologie

Les travaux de l'équipe Condillac reposent principalement sur les notions d'ontologie et de terminologie. Ils ont introduit de nouveaux paradigmes dont l'ontoterminologie: « une ontoterminologie est une terminologie dont le système notionnel est une ontologie formelle et qui intègre des principes issus de la terminologie, de la linguistique, de la logique et de l'intelligence artificielle » (Roche 2007, 2008).

Tout en tenant compte de ces recherches, nous élaborons une ontoterminologie qui part de l'harmonisation conceptuelle et terminologique du domaine de l'éducation médicale. Cet outil destiné à des professionnels de la santé facilitera l'auto-apprentissage autonome des professionnels, ayant des fonctions éducatives particulières, et de l'intercommunication scientifique.

La recherche sur les ontologies se caractérise aujourd'hui par deux grandes tendances: les ontologies formelles et les ontologies élaborées à partir de corpus.

L'ontologie, objet de notre recherche, est organisée à partir de corpus écrits et oraux. Les premiers sont constitués par des textes spécialisés d'œuvres de référence sur l'éducation médicale. Les seconds intègrent des documents enregistrés pendant des congrès et/ou des interviews de spécialistes. Ces deux types de corpus ont comme objectif la collecte de concepts et de termes récents. Pour l'extraction des données linguistiques, nous utilisons des hypertextes et un extracteur de termes.

Dans un premier temps, on recherche les principaux concepts et les différents types de relations qui les unissent pour construire un réseau conceptuel.

Les relations génériques ont un rôle primordial, servant de support au réseau conceptuel; pour l'acquisition de ce type de relation, nous utilisons l'analyse comparative de structures syntaxiques des unités multilexémiques et des marqueurs de relations génériques.

À titre d'exemple, nous présentons, dans la figure 1, les relations du concept *educação médica* (éducation médicale) qui a comme objet *ensino* (enseignement), *aprendizagem* (apprentissage) et *avaliação* (évaluation).

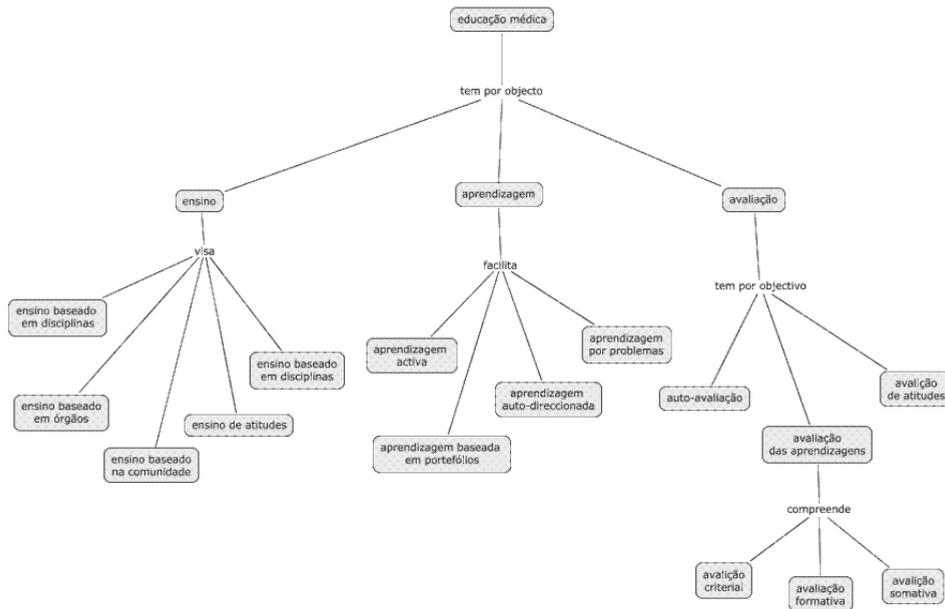


Figure 1

Les relations partitives sont acquises à travers des marqueurs spécifiques; elles peuvent aussi être attribuées automatiquement par inférence. Les relations fonctionnelles sont délimitées par l'analyse de la structure interne des termes complexes et/ou par la cooccurrence des termes dans le même contexte. Les relations spécifiques du domaine peuvent être de différents types.

3 Ontologie et base de connaissances

Une ontologie contient la structure conceptuelle d'une base de connaissances; il existe un consensus pour délimiter précisément les frontières entre ontologie et base de connaissances. Une base de connaissances est plus efficace, en termes de rigueur, que l'ontologie qui lui sert de support, dans la mesure où elle représente des connaissances plus spécifiques. Une base de connaissances est souvent une extension d'une ontologie qui spécifie les concepts fondamentaux du domaine et leurs relations.

Le concept de base de connaissances terminologique BCT (*Terminological Knowledge Base*, I. Meyer et Skuce, projet COGNITERM, 1992) doit spécifier le savoir d'un natif parlant, et, en même temps, du spécialiste du domaine. Une BCT contient une base de données terminologique. Une BCT présente une information linguistique qui sert à constituer une modélisation dans la construction des ontologies formelles/bases de connaissances. Une BCT entretient une relation réciproque avec le corpus-base; Charlet (2007 : 70) affirme qu'« une BCT est attachée à un corpus donné. Elle est le résultat d'un processus d'analyse linguistique méthodique de ce

corpus qui conduit à identifier un réseau de concepts et un ensemble de termes reliés d'une part à ces concepts et d'autre part aux textes, le tout constituant une modélisation (aussi) objective (que possible) du corpus ».

À son tour, Condamines (2003 : 50) considère que les bases de données terminologiques présentent la représentation relationnelle des concepts; les réseaux conceptuels définissent les concepts explicitant leurs interrelations et la terminologie valide la notion de système conceptuel.

En terminologie textuelle, le concept se comprend comme un « signifié normé » (Rastier *et al.* 1994); ce qui signifie que de la relation entre le signifiant et le concept découle l'identification des régularités des usages (normaison) qui présuppose la normalisation.

Selon Costa et Silva (2008 : 8), « parler de terminologie textuelle est plus une question de méthodologie que de théorie; une méthodologie de travail qui aurait pour base les méthodologies propres aux linguistiques de corpus, mais appliquées aux textes de spécialités. Ce qui est [...] spécifique à la terminologie textuelle, c'est le regard porté sur les textes qui, en terminologie, présupposent une approche extralinguistique importante ».

Une base de connaissances terminologique avec son nucléon, l'ontologie, peut être informelle, tout en dépendant de ses applications. Le plus important consiste dans la construction de la représentation du contenu des textes à travers leur description linguistique qui prend en compte de deux opérations : d'une part l'acquisition des termes du domaine, d'autre part leurs relations.

Une ontologie terminologique, « ensemble de concepts d'un domaine structuré par des relations » (Frath *et al.* 2000 : 292), constitue des nucléons de BCT, complètement intégrés ou constituants modulaires, inscrits dans les normes ISO, surtout en ce qui concerne les normes de « concept » et de « système conceptuel » (*cf.* ISO 704, 2000 et ISO 1087, 2001).

4 Ontoterminologie de l'éducation médicale comme outil de formation pédagogique

L'élaboration de cette ontoterminologie se fixe pour but final des modules d'enseignement de l'éducation médicale, disponibles dans une plateforme d'e-learning, qui serviront de support à une formation pédagogique à distance.

Il s'agit de construire un outil à partir d'un corpus de base constitué par de grandes quantités de données textuelles sur support électronique; l'objectif de l'analyse de corpus sera surtout l'acquisition de connaissances lexicales et sémantiques.

Cet outil intégrera des modules de formation relatifs aux différentes tâches pédagogiques des médecins-enseignants des Facultés de médecine et des enseignants d'autres institutions supérieures de sciences de la santé qui veulent développer une autoformation, visant l'appropriation de la compréhension des concepts de l'éducation médicale dans une perspective de réseaux conceptuels qui, d'un point de vue cognitif, se construisent progressivement. On propose aux médecins-enseignants une ontologie du domaine de l'éducation : cela implique l'acquisition de connaissances en sciences de l'éducation dans ce domaine précis.

Les corpus textuels sont analysés de façon à établir la modélisation du domaine. Le logiciel, support de l'ontoterminologie, intègre les différentes relations qui existent entre concepts et qui contribuent à leurs définitions. La construction de cette ontologie définit une modélisation de caractère conceptuel, non ambiguë, qui décrit les connaissances propres à un domaine donné.

Dans une perspective d'ontologie différentielle (Bachimont 2000), la hiérarchie de concepts et de relations est organisée selon les similarités et les différences qu'ils entretiennent les uns avec les autres. Le corpus permet la vérification de la stabilisation sémantique du signifié linguistique pour dégager les objets du domaine et leur type. L'engagement ontologique se construit à partir des corpus textuels du domaine, en dégageant un ensemble de termes et en organisant un système de différences entre ces termes et, finalement, l'opérationnalisation dans un langage de représentation des connaissances (Charlet 2007).

Conclusion

Cette ontoterminologie en langue portugaise présentera des cas de variation ; cependant, elle sera élaborée à partir d'une harmonisation conceptuelle et terminologique.

En effet, l'autonomie scientifique et pédagogique de l'enseignement supérieur en médecine et en sciences de la santé exige des médecins-enseignants une double compétence : maîtrise scientifique de leur spécialité et capacité à prendre des décisions à propos de ce qu'ils enseignent ainsi que de la méthodologie utilisée.

Bref, cette ontoterminologie sera un outil d'aide au travail scientifique et pédagogique en sciences de la santé, en particulier, à celui de l'éducation médicale.

Bibliographie

- Bachimont (B.), 2000 : « Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances », dans Charlet, (J.), Zacklad (M.), Kassel (G.) et Bourigault (D.), éd., *Ingénierie des connaissances : évolutions récentes et nouveaux défis*, Paris, Eyrolles, p. 305-323.
- Bento (J.), 2007 : *Da constituição do corpus à construção de uma ontologia e base de conhecimentos terminológicos*, Tese de Doutoramento, Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas.

- Bratsas (C.), Kaimakamis (E.), Koutkias (V.), Bamidis (P.D.) et Maglaveras (N), 2007 : « An Ontology-Based Approach to Constructing Medical Computational Problems for Use in Electronic Medical Education », dans *Computer-Based Medical Systems*, 2007, CBMS apos ; 07, Twentieth IEEE International Symposium on Volume Issue, 20-22, p. 669-674 <http://users.bestweb.net/~sowa/ontology/ontoshar.htm>.
- Charlet (J.), 2007 : *Ontologies et thésaurus en médecine*, INSERM, UMRS_729, AP-HP, (Avec la collaboration de P. Zweigenbaum ; CNRS, Limsi ; Inalco, CRIM), CNAM – Paris, Bookmarks.
- Charlet (J.), Bachimont (B.) et Jaulent (M.-C.), 2006 : « Building medical ontologies by terminology extraction from texts: An experiment for the intensive care units », dans *Computer in Biology and Medicine*, July-August 2006, 36 (7-8), p. 857-870.
- Condamines (A.), 2003 : *Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques*, Mémoire d'habilitation à diriger les recherches, Toulouse, Université Toulouse 2.
- Contente (M.), 2008 : *Terminocriatividade, Sinonímia e Equivalência Interlinguística em Medicina*, Lisboa, Edições Colibri.
- Costa (R.) et Silva (R.), 2008 : « De la typologie à l'ontologie de textes », dans *Terminologies & Ontologies : théories et applications. Actes de la 2^e Conférence TOTh Annecy 2008*. Annecy, Institut Porphyre. Savoir et Connaissance [à paraître].
- Depecker (L.), 2007 : « Linguistique, terminologie et ontologie », dans Depecker, (L.), Dubois, (V.) et Roche, (Chr.), éd., *Terminologie et ontologie : descriptions du réel, Actes du colloque du 1^{er} décembre 2006*, Paris, Le savoir des mots n° 4, Société française de terminologie, p. 13-18.
- Frath (P.), Oueslati (R.) et Rousselot (F.), 2000 : « Identification de relations sémantiques par repérage et analyse de cooccurrences de signes linguistiques », dans Charlet (J.), Zacklad (M.), Kassel (G.) et Bourrigault (D.), éd., *Ingénierie des connaissances : évolutions récentes et nouveaux défis*, Paris, Eyrolles, p. 291- 304.
- Fürst (F.), 2004 : « Opérationnalisation des ontologies : une méthode et un outil », dans *Ingénierie des Connaissances (IC)*, Lyon, France, p. 199-210.
- Gruber (T.), 1995 : « Toward principles for the design of ontologies used for knowledge sharing », *International Journal of Human-Computer Studies*, vol. 43, n° 5/6, p. 907-928.
- Guarino (N.), 1997 : « Understanding, building, and using ontologies », dans *International Journal of Human-Computer Studies*, vol. 46, n° 2, p. 293-310.
- Guarino (N.), 1998 : « Formal ontology and information systems », dans *Proceedings of Formal Ontologies in Information Systems, FOIS-98*, Trento, Frontiers in Artificial Intelligence and Applications, vol. 46, Amsterdam, IOS Press, p. 3-15.
- ISO 704, 2000 : *Travail terminologique – Principes et méthodes* (français et anglais), Afnor, 2^e édition, 15 novembre 2000, 41 p.
- ISO 1087, 2001, *Travaux terminologiques – Vocabulaire, Partie 1: Théorie et application*, (français et anglais), Afnor, 2^e édition, février 2001, 42 p.
- Rastier (F.) et al., 1994 : *Sémantique pour l'analyse : de la linguistique à l'informatique*, Paris, Masson.
- Roche (C.), 2008 : « Faut-il revisiter les principes terminologiques ? », dans *TOTH 2008 Terminologie et Ontologie : théories et applications*, Annecy, 5 et 6 juin 2008, p. 53-72.
- Roche (C.), 2008 : « Terme et concept : fondements pour une ontoterminologie », dans *TOTH 2007*, Annecy, 1^{er} juin 2007, p. 1-22.

- Smith (B.), 1998 : « An Introduction to Ontology », dans Peuquet (D.), Smith, (B.) et Brogaard (B.), éd., dans *Report of a specialist meeting held under the auspices of varenius project: Panel on computational implementations of geographic concepts*, Bar Harbour, Maine, p. 10-15.
- Sowa (J. F.), 1994 : *Ontology*, www.jfsowa.com/ontology/, date de consultation : 17 juin 2009.
- Sowa (J. F.), 2000 : *Knowledge representation: Logical, philosophical and computational foundations*, New York, Brooks Cole Publishing.
- Sowa (J. F.), 2006 : « A Dynamic Theory of Ontology », dans Bennett (B.) et Fellbaum (C.), éd., *Formal Ontology in Information Systems*, IOS Press, Amsterdam, 2006, p. 204-221.
- Ukor (R.) et Armit (G.), 2008 : *An Ontology of the Organisation and Management of Medical Education*, Manchester Medical School, The University of Manchester.
- Zweigenbaum (P.) et Courtois (P.), 1998 : « Acquisition of lexical resources from SNOMED for medical language processing », dans *Proceedings of the 9th World Congress on Medical Informatics*, Cesnik, Branko, Safran, Charles, Degoulet, Patrice éd., Seoul, p. 586-590.
- Zweigenbaum (P.) et Charlet (J.), (2007) : « Ressources termino-ontologiques pour le domaine médical », dans *Terminologie et ontologie: descriptions du réel, Actes du colloque 1^{er} décembre 2006*, Société française de terminologie, Le savoir des mots, Paris, p. 83-111.

*Marie Calberg-Challot,
Christophe Roche,
Luc Damas,*
Équipe Condillac, Laboratoire Listic,
Université de Savoie, Campus scientifique
73376 Le Bourget du Lac CEDEX,
France,
marie.calberg-challot@orange.fr,
christophe.roche@univ-savoie.fr,
luc.damas@univ-savoie.fr.

Définition du terme vs définition du concept : pour une diversité langagière et une conceptualisation commune

Résumé

Les individus sont de plus en plus amenés à collaborer et à échanger de l'information. Il n'en reste pas moins qu'ils doivent s'accorder sur le sens des mots. La première étape est donc de définir la terminologie d'un domaine. Elle peut nécessiter un travail long et difficile même pour les domaines de spécialité pour lesquels il existe des normes. S'il est souvent difficile de trouver un consensus sur la définition des termes en langue naturelle, l'expérience nous montre qu'il n'en est pas de même pour la conceptualisation formelle d'un domaine où un consensus entre les experts peut être plus facilement atteint. Partant de cet état de fait, nous avons introduit un nouveau paradigme, celui de l'ontoterminologie. L'idée principale est de séparer la dimension linguistique et la dimension conceptuelle de la terminologie. La construction préalable de la partie conceptuelle de la terminologie, socle stable et consensuel pour les experts permet alors de préserver la diversité langagière des communautés de travail. Nous présenterons les résultats de ce nouveau paradigme au travers du dictionnaire métier appliqué au domaine de l'ingénierie hydraulique.

Mots-clés: dictionnaire; terme; concept; terminologie; ontologie; ontoterminologie.

Introduction

Construire une terminologie est une tâche difficile et écrire des définitions qui prennent en compte les différents sens des termes soulève des difficultés même dans des domaines scientifiques et techniques bien circonscrits. Si les experts s'accordent sur une conceptualisation commune, il n'en est pas de même pour les définitions en langue naturelle où le « consensus est toujours relatif à un moment donné, car la variation est toujours possible » (Calberg *et al.* 2008a: 130), mais nous n'aborderons pas dans cet article le problème de la diachronie. Retenons qu'un terme est une « unité linguistique spécialisée » qui dénote un concept du domaine, c'est « le nom

donné dans une langue à une entité conceptualisée par une communauté de travail » (Lerat 2009a : 217) et que le concept, « ensemble unique de caractères » (ISO 1087-1) est un « contenu de connaissance normé et nommé de façon consensuelle » (Lerat 2009b : 76) correspondant à une « représentation intellectuelle permettant de viser le réel suivant des déterminations abstraites et générales et non dans sa singularité concrète » (Baraquin *et al.*). L'expérience nous montre également que dans la construction d'une terminologie, les experts se réfèrent davantage aux schémas, formules ou diagrammes techniques qu'aux textes (Thoiron *et al.* 1996). Cela montre que les experts s'accordent sur un même extralinguistique et sur la définition des concepts quand ceux-ci sont écrits dans un langage semi-formel ou formel. Cela rend leur interprétation objective grâce à un système de règles formelles. L'objectif de cet article est de montrer l'importance de séparer les deux dimensions de la terminologie, à savoir la dimension linguistique et la dimension conceptuelle. Cette séparation des deux systèmes sémiotiques, au-delà de l'opérationnalisation des terminologies, permet d'atteindre un consensus entre les experts tout en préservant la diversité langagière des communautés de travail. La séparation du système linguistique et du système conceptuel entraîne la séparation des définitions, car elles ne relèvent pas du même système sémiotique. Alors que le premier système donnera des explications linguistiques, le second donnera les spécifications logiques. Le résultat est ce qu'on appelle une ontoterminologie et elle sera illustrée au travers du dictionnaire métier.

1 La notion d'ontoterminologie

Séparer la dimension linguistique et la dimension conceptuelle de la terminologie nous amène à introduire le paradigme de l'ontoterminologie (Roche 2007a). L'ontoterminologie, « terminologies dont le système notionnel est une ontologie formelle, insiste sur l'importance des principes épistémologiques qui président à la conceptualisation du modèle – c'est l'ontologie dans sa définition première. Elle insiste également sur la nécessité d'une approche scientifique de la terminologie où l'expert joue un rôle fondamental – c'est l'ontologie dans ses définitions plus récentes où la logique et les langages de représentation des connaissances tiennent une place prépondérante. Enfin, elle met en relation le modèle conceptuel et les termes (d'usage ou normés) qui en parlent, tout en distinguant les définitions formelles des concepts (spécifications logiques) des définitions en langue naturelle des termes (explications linguistiques) » (Roche 2007a, 2009).

D'un côté, la signification du terme est le concept dénoté et l'objectif principal de la terminologie n'est pas de représenter les concepts en vue de leur manipulation comme en intelligence artificielle, mais de définir un vocabulaire commun et consensuel. En terminologie, le concept n'existe pas pour lui-même, mais à travers sa définition en langue naturelle.

D'un autre côté, la conceptualisation est essentielle dans un domaine de spécialité. Elle est construite selon une théorie donnée en utilisant un langage semi-formel ou formel en suivant des principes épistémologiques. Les spécifications logiques du

concept sont identifiées et les experts s'accordent sur la dénomination du concept pour lever les ambiguïtés.

1.1 Dire n'est pas concevoir

L'ontoterminologie réunit deux systèmes de signes qui sont distincts l'un de l'autre. Le système linguistique est directement lié aux textes et aux discours de spécialité tandis que le système conceptuel concerne la modélisation du domaine. Il en découle naturellement que la rédaction de documents dans un domaine de spécialité n'est pas la même activité que celle de la modélisation des concepts. Même si l'extraction de termes candidats est une source importante dans la construction d'une terminologie (Buitelaar *et al.* 2005; Daille *et al.* 2004; Calberg-Challot *et al.* 2008b), dire n'est pas concevoir (Roche 2007b) et la structure lexicale ne se superpose pas à la structure conceptuelle (Rastier 1995, 2004; Roche 2007b; Calberg-Challot *et al.* 2008a; Desprès et Szulman 2009). Il est donc essentiel de définir distinctement ces deux activités complémentaires dans la construction de l'ontoterminologie.

1.2 Terme et Concept

L'ontoterminologie permet de visualiser la terminologie par les deux systèmes sémiotiques, c'est-à-dire par le système des termes aussi bien que par le système des concepts. Ce nouveau paradigme est le cœur du dictionnaire métier que nous allons présenter dans la troisième partie. Cela signifie que les définitions des termes et des concepts sont séparées, mais également que les deux systèmes sont en étroite relation. L'ontoterminologie permet grâce à deux systèmes distincts de visualiser les deux dimensions de la terminologie. Si l'ontoterminologie suit une certaine normalisation, l'objectif est de préserver et de servir la diversité langagière et il est nécessaire de valider des définitions par le genre et la différence en les reliant à des définitions encyclopédiques qui donnent des descriptions du terme (Lerat 2009a).

2 Une méthodologie pour la construction du dictionnaire métier

L'objectif est la définition de terminologies qui soient consensuelles, partageables et cohérentes pour plusieurs communautés de travail. Pour cela, nous adoptons une démarche onomasiologique dans la construction du dictionnaire métier et les terminologies reposent sur la notion d'ontoterminologie.

La modélisation d'un domaine et la rédaction de discours à laquelle elle peut donner lieu sont deux activités différentes. La modélisation construite directement par les experts à l'aide d'un langage formel et la modélisation construite à partir de textes ne se superposent pas (Roche 2007b). De plus, une modélisation construite à partir de documents est dépendante du corpus et ne vérifie pas l'ensemble des propriétés recherchées.

C'est pourquoi nous mettons l'accent sur la conceptualisation du domaine et la présence des experts est alors indispensable.

La méthode se décompose en cinq étapes dont voici le schéma de réalisation (figure 1).

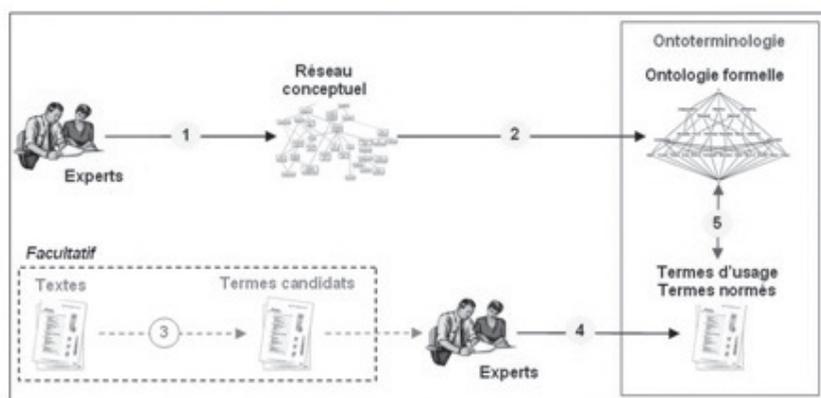


Figure 1

La modélisation semi-formelle du domaine a pour objectif d'identifier, avec l'aide des experts, les concepts du domaine et les différentes relations qui les lient (relations de subsomption de type « est un », relations de spécialisation de type « est une sorte de », relations de composition de type « est composé de »...). Le résultat est un réseau conceptuel. La construction du réseau conceptuel est réalisée en collaboration avec les experts à l'aide de l'environnement iMap de construction de réseaux conceptuels.

Les concepts ayant été identifiés, nous pouvons les organiser sous la forme d'une ontologie formelle, c'est-à-dire sous la forme d'un arbre de concepts liés par la relation de généralisation-spécialisation. Les concepts sont définis à l'aide de l'environnement iConcept de construction d'ontologies par différenciation spécifique (Roche 2001).

Afin d'aider les experts dans le choix des termes d'usage, c'est-à-dire des termes métier utilisés dans les discours, qu'ils soient oraux ou écrits, des termes candidats sont générés automatiquement à partir d'un corpus de référence à l'aide de l'outil iWord. Notons que cette étape est facultative et qu'elle intervient dans un second temps, une fois la conceptualisation du domaine réalisée. La construction de la terminologie pourrait commencer par la génération de termes candidats afin d'initier le processus en identifiant certains concepts. Cependant, dans un souci de fournir un socle stable aux experts et en insistant sur l'importance de séparer les deux systèmes sémiotiques, nous introduisons l'étude des termes candidats une fois la conceptualisation réalisée.

Les experts identifient, lors de la quatrième étape, les termes d'usage et les termes normés. Ils pourront se servir des termes candidats générés lors de la phase précédente. Les termes d'usage sont les termes utilisés en discours (oraux ou écrits). Ils permettent de prendre en compte la diversité langagière. Ils peuvent être polysémiques. Les termes normés sont des termes univoques qui sont soit en usage soit prescriptifs.

La dernière étape consiste à mettre en regard, c'est-à-dire à associer, les termes d'usage et les termes normés avec les concepts de l'ontologie correspondants. Les résultats sont ensuite intégrés dans le dictionnaire métier dont nous allons exposer les fonctionnalités.

3 Présentation du dictionnaire métier

Cette partie est consacrée à la présentation du dictionnaire et à ses diverses fonctionnalités. Ce dictionnaire d'un genre nouveau a été l'objet d'un travail pluridisciplinaire où linguistes, cognitivistes, logiciens et informaticiens ont travaillé de concert. Des retours sur l'utilisation ont également eu lieu par divers experts dans plusieurs domaines de travail. Cette présentation sera illustrée dans le domaine de l'ingénierie hydraulique.

3.1 Lancement du dictionnaire métier et choix du domaine

L'interface de lancement du dictionnaire permet de choisir le domaine sur lequel porte une recherche. L'interface du dictionnaire métier se découpe en trois zones (voir la figure 2, ci-après). La première zone [1] permet de visualiser le dictionnaire soit par la liste des concepts (à partir des relations de composition ou des relations de spécialisation), soit par la liste des termes classés par ordre alphabétique. La deuxième zone [2] permet de visualiser la conceptualisation à travers des relations de spécialisation ou de composition. Enfin, la troisième zone [3] présente deux autres modes de visualisation du dictionnaire : le mode de visualisation par hyperschéma ou le mode de visualisation par la fiche terminologique. Dans ce deuxième cas, deux fiches terminologiques sont alors accessibles : les fiches terminologiques conceptuelles et les fiches terminologiques linguistiques.

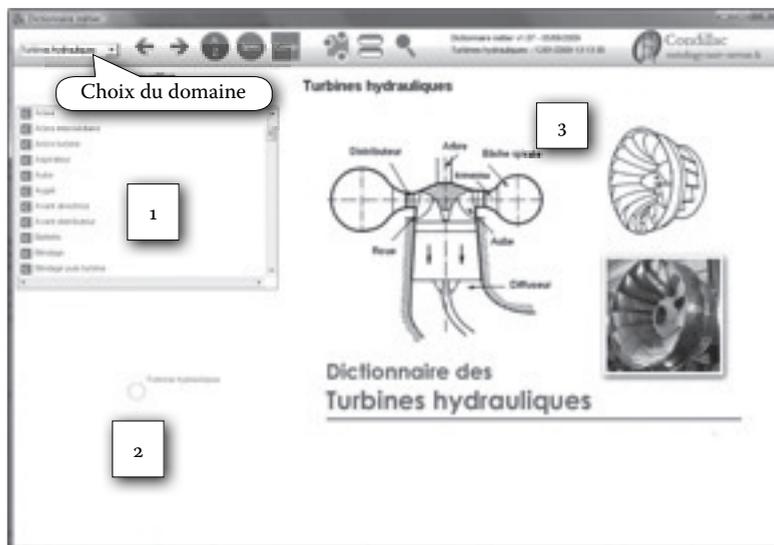


Figure 2

3.2 Principales fonctionnalités du dictionnaire métier

Le choix d'un domaine charge le dictionnaire demandé. Une fois le dictionnaire lancé, plusieurs types de visualisations sont possibles. Il est possible de naviguer à travers les concepts ou les termes. Dans la figure 3, ci-après, le choix a été fait sur une navigation par concepts reliés par une relation de composition. L'exemple ci-après montre qu'un < Groupe hydraulique Francis > est composé d'une < Turbine hydraulique Francis > et d'un < Alternateur > et correspond à la définition formelle du concept. Les chevrons mettent l'accent sur le système conceptuel. L'interface permet de visualiser en même temps où se situent les éléments de composition sur l'hyperschéma. Ce mode de visualisation est important, car « il y a [...] quelque chose d'éternel dans un schème technique » (Simondon 1989) et « la représentation graphique d'un objet est souvent irremplaçable » (Lerat 1995 : 149).

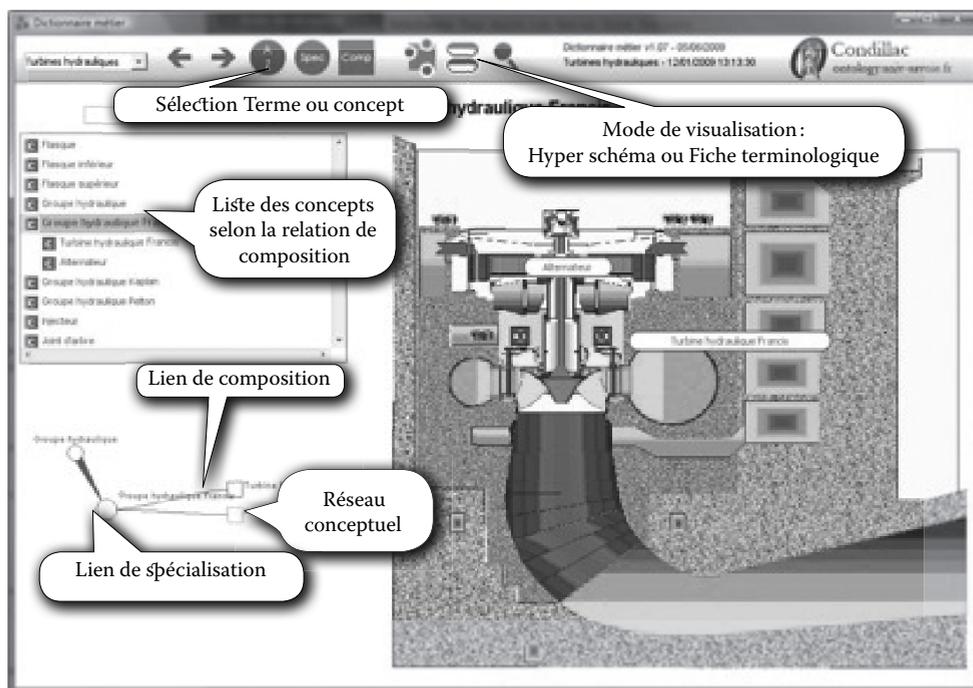


Figure 3

3.3 Navigation par les hyperschémas, réseaux et liste de concepts

Les relations entre les concepts sont utilisées pour structurer les entrées du dictionnaire. Les relations sont des relations de composition de type « est composé de » ou de spécialisation de type « se spécialise en ». Chaque utilisateur peut choisir la visualisation qui lui semble la plus pertinente selon son besoin.

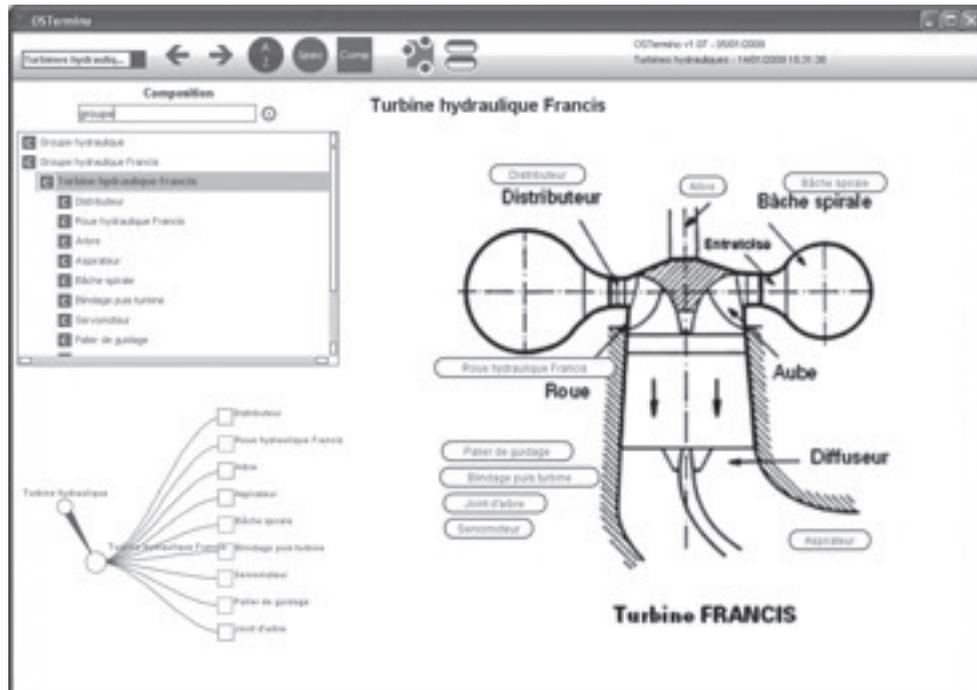


Figure 4

3.4 Navigation par les fiches terminologiques

Le dictionnaire métier fournit deux types de fiches : une fiche conceptuelle et une fiche terminologique où les définitions apparaissent distinctement. La fiche du haut est consacrée au concept et à sa définition formelle. Pour bien démarquer les deux systèmes sémiotiques, une convention typographique a été adoptée. Ainsi, les concepts sont encadrés par des chevrons ouvrants et fermants. Dans cette fiche terminologique conceptuelle, on trouve la définition formelle où, dans la figure ci-après, une < Turbine hydraulique Francis > est composée d'un < Distributeur >, d'une < Roue hydraulique Francis >, d'un < Arbre >, d'un < Aspirateur >, d'une < Bâche spirale >, d'un < Blindage puits turbine >, de < Servomoteur >, d'un < Palier de guidage > et d'un < Joint d'arbre >. Cette fiche indique les termes normés et les termes d'usage qui désignent le concept auquel ils se rapportent.

La fiche du bas, quant à elle, est consacrée au terme et à sa définition en langue naturelle. C'est dans cette fiche que toutes les informations en langue telles que la synonymie, l'hyponymie, l'hyperonymie et les équivalents étrangers apparaissent. Les termes sont mentionnés entre guillemets. Ainsi, la « turbine hydraulique Francis » sera définie comme « machine qui transforme l'énergie hydraulique en énergie mécanique. Le terme n'inclut ni la vanne d'entrée ou de sortie ni le générateur ou le régulateur associé (Source Norme CEI-TR 61 364) ».

3.5 Navigation classique par les articles du dictionnaire

Cette dernière partie présente la navigation dans le dictionnaire métier à travers les termes. La fiche terminologique du bas du dictionnaire métier est alors consacrée au terme d'usage (figure 5).

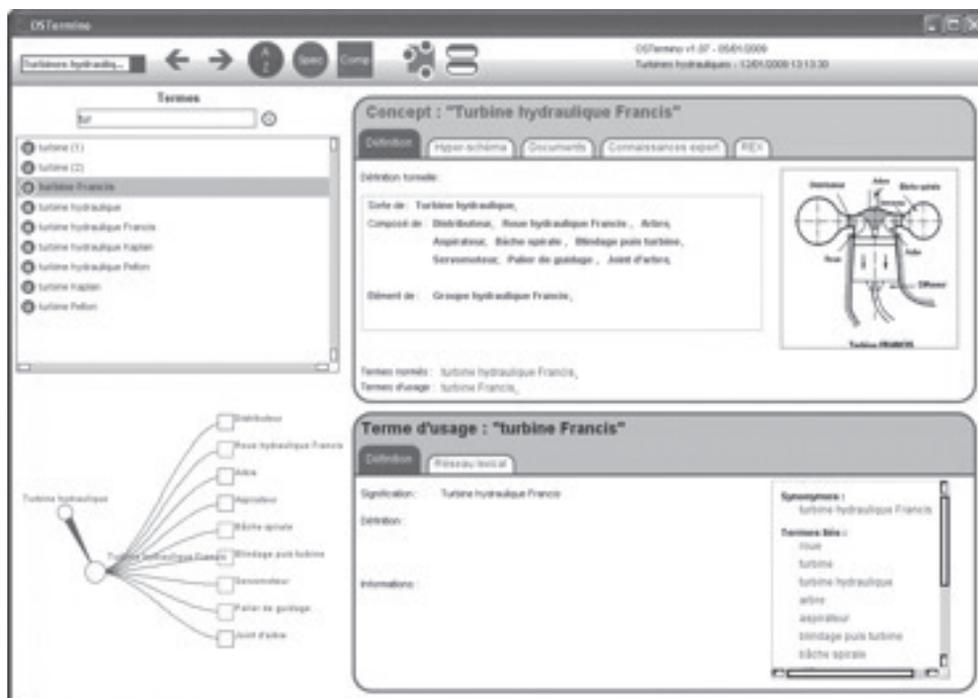


Figure 5

4 Dictionnaire au service de la diversité langagière

Dans cette dernière partie, nous souhaitons montrer que la conceptualisation formelle est bien au service de la diversité langagière. En effet, toute personne cherchant des informations sur un terme le fera sans avoir à modifier ses usages. Le dictionnaire métier donne à l'utilisateur la possibilité de conserver sa terminologie pour faire ses recherches. L'ensemble des termes d'usage ou normés et le concept auxquels les termes se rapportent seront retournés à l'utilisateur. Tout en gardant son usage ou le terme trouvé dans un document, le système retourne l'information recherchée. Dans la figure ci-après, l'utilisateur cherchant des informations sur le terme « queue d'aronde » trouvera les autres usages tels « queue d'ironde » ou « clé d'ironde ». Le dictionnaire métier le renseignera en même temps sur le concept < Dispositif d'accrochage par clés > et le terme normé « Dispositif d'accrochage par clés ». L'utilisateur, en pouvant télécharger l'ontologie du domaine, la définition formelle et la définition en langue naturelle ne pourra pas se tromper sur la signification du terme et connaîtra l'ensemble des différents dispositifs d'accrochage.

The image displays three overlapping screenshots of a software interface for ontology management, likely using the OK Grapher tool.

- Top Window:** Shows a concept definition for "Dispositif d'accrochage par clés d'aronde". It includes a title bar, navigation buttons, and a text area for the definition. A small image of a mechanical part is visible on the right.
- Middle Window:** Displays a hierarchical graph of the ontology. The root node is "Dispositif d'accrochage", which branches into "par clés" and "par goujons". "par clés" further branches into "en T" and "aronde". "par clés en T" and "Dispositif d'accrochage par clés d'aronde" are shown as sub-nodes.
- Bottom Window:** Shows a detailed view of the concept "Dispositif d'accrochage par clés". It includes a title bar, navigation buttons, and a text area for the definition. A "Voir" button is present. Below the definition, there is a section for "Termes normés" and "Termes génériques" related to the concept.

Conclusion

Les experts attendent une terminologie qui définit clairement les termes portant sur la conceptualisation de leur domaine. Même si les définitions en langage naturel sont toujours utiles, les experts éprouvent certaines difficultés à trouver un consensus; ce qui n'est pas le cas pour la conceptualisation où les experts s'accordent sur la dénomination du concept tout en préservant les divers usages qui le désignent. Les experts sont soucieux de préserver la diversité langagière, qui est importante car elle marque une identité professionnelle, et ils valident le fait

que plusieurs termes puissent dénoter un même concept. Nous avons introduit le paradigme d'ontoterminologie, terminologie dont le modèle conceptuel est une ontologie formelle, dans l'objectif de séparer la définition du terme de la définition du concept. Ceci implique que le concept n'est pas un terme, ni la définition du terme. Ce nouveau paradigme, au travers d'une conceptualisation commune, permet une meilleure compréhension entre les communautés de travail tout en préservant la diversité langagière. Une telle approche ouvre de nouvelles voies de travail, en particulier pour les traducteurs, où les savoirs peuvent être rattachés à l'ontoterminologie. Enfin, ce dictionnaire métier ne cessera de s'enrichir et de s'améliorer au fur et à mesure du retour des utilisateurs.

Bibliographie

- Baraquin (N.), Baudart (A.), Dugué (J.), Laffitte (J.), Ribes (Fr.), Wilfert (J.), (1995) : *Dictionnaire de philosophie*, Paris, Armand Colin, Collection Dictionnaire, 384 p.
- Buitelaar (P.), Cimiano (P.), Magnini (B.), 2005 : « *Ontology Learning from text : Methods, Evaluation and Applications* », dans *Frontiers in Artificial Intelligence and Applications*, Vol. 123, P. Buitelaar Editor, Ios Press Publication, July 1.
- Calberg-Challot (M.), Candel (D.), Roche (Chr.), 2008a : « De la variation des usages au consensus terminologique : vers un dictionnaire de l'ingénierie nucléaire », dans *Actes de la première conférence TOTH 2007, Terminologie & Ontologie : Théories et applications*, Roche (Chr.), éd., Annecy, Institut Porphyre, p. 199-141.
- Calberg-Challot (M.), Candel (D.), Bourigault (D.), Dumont (X.), Humbley (J.), Joseph (J.), 2008b : « Une analyse méthodique pour l'extraction terminologique dans le domaine du nucléaire », dans *Terminology*, 14:2, p. 183 – 203.
- Daille (B.), Kageura (K.), Nakagawa (H.), Chien (L. F.), 2004 : « Recent Trends in Computational Terminology », dans *Special issue of Terminology*, 10 :1, Benjamins publishing company.
- Desprès (S.), Szulman (S.), 2009 : « Réseau terminologique versus Ontologie », dans *Actes de la deuxième conférence TOTH 2008, Terminologie & Ontologie : Théories et applications*, Roche (Chr.), éd., Annecy, Institut Porphyre, p. 17-34.
- ISO 1 087-1, 2001. Vocabulaire, ISSN 0335-3931.
- Lerat (P.), 1995 : *Les langues spécialisées*, Paris, PUF.
- Lerat (P.), 2009a : « La combinatoire des termes. Exemple: Nectar de fruits », dans *Hermes Journal of Language and Communication Studies*, 42, p. 211-232.
- Lerat (P.), 2009b : « Propositions pour un réseau conceptuel des instruments de mesure œnologiques », dans *Actes de la deuxième conférence TOTH 2008, Terminologie & Ontologie : Théories et applications*, Roche (Chr.), éd., Annecy, Institut Porphyre, p. 73-90.
- Lino (T.), 2006 : « Contextes et néologie terminologique dans le domaine médical », dans *Mots, termes et contextes*, Blampain (D.), Thoiron (Ph.), Van Campenhoudt (M.), éd., Paris AUF, p. 509-514.
- Rastier (Fr.), 1995 : « Le terme : entre ontologie et linguistique », dans *La banque des mots*, numéro spécial 7-1995, p. 35-65.
- Rastier (Fr.) 2004 : « Ontologie(s) », dans *Revue d'Intelligence Artificielle*, vol. 18 n° 1, p. 15-40.

- Roche (Chr.), 2001 : « The « Specific-Difference » Principle: a Methodology for Building Consensual and Coherent Ontologies », dans *Actes de la conférence IC-AI'2001*, Las Vegas, USA.
- Roche (Chr.), 2005 : « Terminologie et ontologie », dans *Langages*, n° 157, p. 48-62.
- Roche Christophe (2007a) : « Le terme et le concept: fondements d'une ontoterminologie », dans *Actes de la première conférence TOTH 2007, Terminologie & Ontologie: Théories et applications*, Roche (Chr.), éd., Annecy, Institut Porphyre, p. 1-22, 2007.
- Roche (Chr.), 2007b : « Dire n'est pas concevoir », dans *IC 2007 : 18^{es} Journées Francophones d'Ingénierie des Connaissances*, Grenoble 2-6 juillet 2007.
- Roche (Chr.), 2009 : « Faut-il revisiter les principes terminologiques ? », dans *Actes de la deuxième conférence TOTH 2008, Terminologie & Ontologie: Théories et applications*, Roche (Chr.), éd., Annecy, Institut Porphyre, p. 53-72.
- Simondon (G.), 1989 : *Du mode d'existence des objets techniques*, Aubier philosophie, (1958, 1969), 3^e édition.
- Thoiron (Ph.), Arnaud (P.), Béjoint (H.), Boisson (Cl. P.), 1996 : « Notion « d'archi-concept » et dénomination », dans *Meta*, vol. 41, n° 4, p. 512-524.

Natalie Kübler,
Geneviève Bordet,
Mojca Pecman,
CLILLAC-ARP EILA,
Université Paris Diderot,
France,
nkubler@eila.univ-paris-diderot.fr,
mpecman@eila.univ-paris-diderot.fr,
gbordet@eila.univ-paris_diderot.fr.

La linguistique de corpus entretient-elle d'étroites relations avec la traduction pragmatique ?

Résumé

Nous considérons les apports de la linguistique de corpus à la traduction pragmatique, démontrant que l'étude des corpus, loin de jouer un rôle de normalisation, ouvre la possibilité d'une créativité linguistique permettant une traduction efficace. Nous en donnons deux exemples : les rapports entre prosodie sémantique et l'expression de l'évaluation dans des articles scientifiques en sciences de la terre, l'analyse comparative de résumés de thèse réalisés en anglais par des locuteurs anglophones et francophones. Nous montrons que la confrontation aux corpus permet de prendre conscience de la diversité des stratégies de communication et d'en envisager le transfert dans la langue cible.

Mots-clés : linguistique de corpus ; traductologie ; langue de spécialité ; traduction pragmatique.

Introduction

On a amplement démontré, depuis plus de dix ans, l'intérêt que pouvait avoir l'utilisation d'un corpus dans le processus de traduction, ainsi que dans son enseignement (Aston 1999 ; Bowker 1998, 2000, Bowker & Pearson 2002, Kübler 2003, Zanettin *et al.* 2003, Bernardini & Castagnoli 2008, Beeby *et al.* 2009). A aussi été abordée la question de l'apport du corpus en traductologie, notamment grâce à l'étude de corpus parallèles alignés (Baker 1999 par exemple). Le corpus est cependant souvent présenté comme un outil permettant aux praticiens, aux enseignants et aux théoriciens de la traduction de trouver des réponses à leurs questionnements. Le traducteur y cherche des équivalents terminologiques ou phraséologiques, des définitions pour mieux comprendre les termes, la bonne collocation dans la langue cible, etc. En traductologie, le corpus permet de mieux étudier les différentes stratégies de traduction, l'influence de celle-ci sur le texte cible, etc. On entend par conséquent « corpus » et non « linguistique de corpus ». Notre objectif

ici consiste donc à éclairer les liens qui se tissent aujourd'hui entre traductologie et linguistique de corpus, dans une optique praxéologique. Dans son étude empirique de la langue, allant au-delà des systèmes formels, la linguistique de corpus rejoint en effet le nécessaire besoin de conscience linguistique du traducteur et suit un chemin parallèle à une approche théorique de la traduction pragmatique. Quelques exemples illustrant des questions liées au genre textuel, à la lexicogrammaire et aux prosodies sémantiques viennent appuyer cet éclairage.

1 Linguistique de corpus et traduction pragmatique : créativité linguistique et normalisation

1.1 Traduction pragmatique, discours spécialisé et genre

Avant d'aborder la question des liens entre linguistique de corpus et traduction pragmatique, il convient de définir d'abord le type de textes auxquels s'attache la traduction spécialisée.

Ces textes appartiennent tous à des domaines de spécialité bien définis tels que le droit, l'ingénierie, la technologie ou les techniques. À l'intérieur de ces domaines, ils se définissent par leur inscription dans un genre spécifique, dont la nature est extrêmement variée : article scientifique, mais aussi résumé d'article, mode d'emploi, publicité, bulletin météo, dépêche d'agence, etc. ; genre qui lui-même est déterminé par un contexte et des objectifs de communication particuliers.

Abordant maintenant la question de la traduction pragmatique, nous pouvons prendre une définition simple de la traduction proposée par Bowker & Pearson (2002), pour qui l'acte de traduire consiste à prendre un texte dans une langue, appelée langue source, et à le recréer dans une autre langue, la langue cible. L'usage du terme « pragmatique » ajoute ainsi la référence à la réception de la traduction par le lecteur (Newmark 1988) et met l'accent sur l'effet perlocutoire de la traduction (Hickey 1998)

La traduction pragmatique, s'attachant au transfert dans une autre langue d'un texte appartenant à un genre spécialisé, doit prendre en compte le type de lectorat auquel elle s'adresse, ainsi que ses attentes.

1.2 Traduire un texte spécialisé

Attachée à traduire un type de textes spécifique, la traduction spécialisée tend également à être définie par l'importance de ses choix terminologiques, visant à trouver les équivalents corrects dans la langue cible. Ces choix terminologiques s'accompagnent nécessairement de choix quant à la manière d'utiliser ces termes en les enchâssant dans une structure, en fonction d'une phraséologie, elle aussi spécialisée. Si l'usage d'une syntaxe correcte est indispensable, ce n'est pas ce qui pose le plus problème au traducteur professionnel (les apprentis par contre montrent encore souvent des difficultés dans le traitement de la syntaxe). En

revanche, il devra recréer ce que l'on désigne couramment comme le « style » particulier au texte. Ainsi, dans le genre mode d'emploi en français, on n'utilise pas les deuxièmes personnes, mais plutôt la troisième personne avec, comme sujet, par exemple, « l'utilisateur », ou l'infinitif. Cette question de style ressortit en fait au genre. En effet, le genre se définit d'une part par des critères externes tels que l'identité de l'auteur, celle de celui à qui il s'adresse, la nature des connaissances, des compétences et des attentes des lecteurs, l'objectif du texte, etc. Le genre présente aussi, d'autre part, des caractéristiques internes qui sont en fait les caractéristiques linguistiques, comme ce qu'on appelle le style, mais aussi des questions marquées linguistiquement et dont on parle moins dans ce domaine, comme les phénomènes de registre, de prosodie sémantique, de positionnement du locuteur/auteur, de l'évaluation ou de l'argumentation.

1.3 Traduire les références scientifiques en français scientifique

Le défaut de prise en compte de ces critères de genre dans les pratiques de traduction explique des décalages parfois importants entre la production de textes en langue cible et en langue source. Ainsi, aujourd'hui encore par exemple, on rencontre chez les étudiants en traduction d'étranges formes de références dans des articles scientifiques traduits en français. En effet, on trouve des éléments comme : « Dans son ouvrage intitulé *Translation and Conflict: A Narrative Account* et publié en 2006, M^{me} Mona Baker démontre ». Une vérification dans un corpus d'articles scientifiques français permet de lever tout doute éventuel sur les modes de citation en vigueur.

En recherchant simplement des dates, on trouve les exemples suivants :

- [1] par le travail de T. Dos Reis (2000) sur le raccord avec la plaine a
- [2] lors du programme SISTEUR (2000) présente des hydrates de gaz,
- [3] En effet, Migeon *et al.* (2000) ont montré que les sédiments
- [4] Sumner et Westbrook (2001) suggèrent que le début de
- [5] Saquaque *et al.* (1992) considèrent le socle comme

1.4 Traductologie

Depuis une quinzaine d'années, les méthodes de la linguistique de corpus sont utilisées en traductologie et dans l'enseignement de la traduction. Mona Baker a été l'instigatrice de cette approche de la traductologie. Elle utilise les corpus et les méthodes statistiques de la linguistique de corpus pour observer et analyser les différentes stratégies des traducteurs qui conduisent ceux-ci à produire un type particulier de textes dans la langue cible, et que Baker caractérise comme appartenant à la variante « *translationese* de la langue ». Pour Baker (1999), la perception qu'ont les traducteurs de la réception sociale d'une traduction, les conduit à se conformer de très près aux caractéristiques typiques de la langue cible et même à les exagérer. Elle remarque par exemple que les traductions contiennent plus de mots déjà habituellement fréquents dans la langue cible que les textes

rédigés directement par des locuteurs natifs dans leur langue; et que la densité lexicale des traductions est plus faible car elle contient plus de mots grammaticaux. La langue des traductions est aussi moins variée. Baker souligne également le danger que peut représenter l'aspect très statistique de la linguistique de corpus: en effet, à force d'étudier des phénomènes linguistiques sous l'angle statistique, on peut avoir tendance à prendre la fréquence comme le seul critère et normaliser à outrance la langue, en oubliant la créativité linguistique qui doit exister aussi en traduction, même en traduction spécialisée. Ce danger nous semble pouvoir être évité si l'on ne s'appuie pas seulement sur des observations statistiques, mais aussi sur les observations linguistiques parfois surprenantes que permet de faire la linguistique de corpus.

1.5 Linguistique de corpus et traduction

Cette conception de la linguistique de corpus permet d'éviter une approche normalisatrice de la langue fondée sur les fréquences d'occurrences de formes, pour lui substituer une démarche d'observation ouverte du discours dans la diversité de ses réalisations. Nous pensons en effet que le développement d'une conscience linguistique aiguë est un élément indispensable au développement d'une compétence de traducteur. Jusqu'ici, on a souvent présenté la linguistique de corpus, soit comme une méthode d'analyse des stratégies du traducteur, soit comme outil permettant d'obtenir des réponses, au même titre que les dictionnaires monolingues et bilingues. Or, la linguistique de corpus ouvre des perspectives sur la langue permettant d'aller beaucoup plus loin, ce qui devrait en faire un élément important de l'éducation d'un traducteur, au même titre que les études de civilisation. Certes, on a produit jusqu'ici, sans s'appuyer sur la linguistique de corpus, des traductions d'excellente qualité. Cependant, les métiers de la traduction ont changé: les outils exigés, la quantité de textes à traduire, leur variété, ont augmenté, alors que le temps mis à la disposition des traducteurs s'est réduit. Utile à une approche pragmatique, la linguistique de corpus permet également de progresser dans l'analyse théorique de la traduction. Nous en prendrons pour exemple l'expression de l'évaluation dans différents genres de discours et sa traduction.

1.6 Traduction et prosodie sémantique

Le fonctionnement de l'évaluation dans le texte est l'une des caractéristiques distinctives du genre. Les avancées réalisées dans ce domaine grâce à la linguistique de corpus ont contribué notamment à mettre en évidence le phénomène de « prosodie sémantique » (Stubbs 1996). On le résumera ici brièvement en disant que certains mots ont tendance à entrer en cooccurrence avec des mots portant une certaine connotation. Pour donner un exemple simple, on considérera ainsi qu'un mot comme *commettre* a une prosodie sémantique négative puisqu'on le retrouve très souvent associé à des mots comme *crime*, *délit*, *meurtre*, *attentat*, etc. De ce fait, l'expression *commettre un roman* sera ressentie comme ironique puisqu'elle associe un terme à connotation négative avec un terme associé aux valeurs positives de création, d'art et d'imaginaire. L'expression « commettre un roman » implique donc un jugement négatif sur le roman et/ou sur son auteur. De manière

moins attendue, on retrouve la prégnance de cette fonction évaluative dans les articles scientifiques, pourtant considérés comme tenus à la neutralité de l'objectivité scientifique.

La connaissance et la compréhension de phénomènes tels que l'évaluation ou la prosodie sémantique permet de les identifier dans un texte source et donc de les traduire correctement dans la langue cible. Mais cela ouvre aussi au traducteur la possibilité de jouer avec la créativité et les variations linguistiques pour produire un effet équivalent. On voit ainsi que, loin d'être un facteur de normalisation, la linguistique de corpus, par la prise de conscience linguistique large et comparative qu'elle permet, offre de nouvelles possibilités de créativité dans le processus de traduction.

1.7 L'évaluation dans l'article scientifique : traduire en jouant sur les variations structurelles

Ainsi, en ce qui concerne les articles scientifiques, l'étude de la fonction d'évaluation à travers un corpus d'articles dans le domaine des sciences de la terre nous permet d'observer, au-delà de la diversité de catégories grammaticales susceptibles d'assurer la fonction d'évaluation, une phraséologie fortement codifiée. Par exemple, les adverbes axiologiques que l'on rencontre auprès du participe « *understood* » dévoilent une double polarité : positive (« *well, clearly understood* ») et négative ou encore, restrictive (« *incompletely, poorly, less well, not well, not fully, not clearly understood* »), tendant plutôt vers le négatif. En français, son équivalent direct, le participe « compris », possède une prosodie sémantique encore plus fortement négative ou restrictive (« assez bien, moins bien, beaucoup moins bien, plutôt bien, mal compris ») et apparaît souvent dans un contexte de négation (« ne sont pas mieux compris », « ne sont toujours pas compris »). En français, l'évaluation positive est plutôt dévolue au participe « connu » et son équivalent nominal (« bien, le mieux, parfaitement connu » ; « bonne, meilleure connaissance », « connaissance approfondie, fine, détaillée, précise »). Les deux admettent cependant également une qualification négative (« peu, faiblement, mal connu » ; « connaissance insuffisante, limitée »).

Cette analyse nous permet ainsi de résoudre la difficulté de traduction que l'on rencontre dans une phrase comme :

[6] *The underlying processes causing Shishaldin's LP events are still poorly understood*

qui pourrait être traduit en français :

[7] Les processus sous-jacents responsables des événements LP de Shishaldin sont encore mal compris

ou plutôt :

[8] Nos connaissances sur les processus sous-jacents responsables des événements Longues-Périodes (LP) de Shishaldin sont encore insuffisantes

La compréhension fine des modes d'expression de l'évaluation dans le genre spécifique de l'article scientifique permet de proposer en langue cible une traduction satisfaisante parce qu'elle joue le même rôle dans un contexte de communication équivalent.

On voit donc bien que la question de traduction s'aborde au-delà du processus de transfert linguistique, dans une appréhension des caractéristiques de la situation de communication liée au genre de discours. La linguistique de corpus, par les perspectives comparatives qu'elle offre, permet de faire avancer la réflexion non seulement sur la traduction interlangue mais aussi pour une approche contrastive monolingue.

Nous en présentons un exemple avec l'analyse contrastive de résumés de thèse réalisés en anglais par des auteurs francophones et anglophones. La comparaison permet de mettre en évidence les stratégies de communication propres à chaque discipline dans la diversité de leurs modes de réalisation, ce dont nous verrons par la suite les implications en termes de traduction.

2 Une approche contrastive monolingue du résumé de thèse

Notre objectif est de contribuer à la compréhension du fonctionnement du discours scientifique et des problèmes que pose sa réalisation en langue 2, ici en anglais, par des locuteurs non natifs, les résumés des locuteurs francophones étant traduits par eux-mêmes en anglais.

2.1 Analyse de discours et linguistique de corpus

Ce travail allie les outils fournis par la linguistique de corpus (Biber 2006, 2008), de la grammaire fonctionnelle (Halliday 1993, 1994; Gledhill 2000) et l'analyse de discours (Hyland 2000) pour la compréhension du fonctionnement de la langue dans un domaine de spécialité.

Dans une perspective d'analyse de discours, le résumé, ou *abstract*, de thèse est considéré comme un genre spécifique du discours scientifique (Swales 1990) structuré par des « mouvements » rhétoriques spécifiques (Biber, Swales).

Ce discours scientifique se caractérise par un fonctionnement syntaxique et sémantique particulier, que peut décrire la « lexicogrammaire » de Halliday.

La linguistique de corpus permettra de vérifier, généraliser et objectiver les hypothèses émises grâce à l'analyse de discours, selon une méthode préconisée par Biber (2008) et Hyland (2000).

2.2 Fonction communicative du résumé de thèse comme genre

Le résumé de thèse est considéré comme un objet autopromotionnel (Hyland 2000) qui fonde sa légitimité sur un réseau intertextuel complexe, en liaison externe

avec ses référents théoriques, interne avec le texte de la thèse. Il a pour double objectif (Wable & Holzem 2004) de prouver au lecteur qu'il doit lire la thèse et de le convaincre que son auteur est un prétendant légitime à l'appartenance à sa communauté scientifique. Pour ce faire, il doit afficher trois qualités essentielles : cohérence de l'argumentation, légitimité du point de vue adopté par la recherche, construction d'une « autorité » d'auteur et de chercheur (« *authorial stance* » chez Hunston & Thompson 1999).

Le résumé de thèse est un objet particulièrement intéressant dans la mesure où il s'agit d'un discours en construction dont les failles sont susceptibles de rendre le fonctionnement plus transparent. Ses enjeux de communication sont très forts, communs aux locuteurs de L1 et L2. Enfin, la situation des thèses françaises est spécifique puisqu'elles restent le seul écrit scientifique dont l'auteur est unique, qu'elles sont rédigées en français, mais accompagnées d'un résumé d'auteur, en anglais et français.

Cependant, nous avons fait le choix, dans un premier temps, de nous concentrer sur une comparaison des résumés de thèses réalisés en anglais par des locuteurs natifs et non natifs, faisant ainsi du corpus français un référent susceptible de fournir des clés d'interprétation des écarts constatés entre les réalisations de résumés en langue 1 et en langue 2. La mise en parallèle de deux corpus réalisés dans la même langue vise à permettre l'évaluation non pas d'une bonne ou d'une mauvaise traduction, mais de stratégies de communication différentes permettant ensuite de revenir sur l'équivalent original français.

2.3 Une démarche d'analyse croisée

Nous avons adopté une double démarche inspirée par Hyland et Biber : analyse des mouvements rhétoriques des textes, puis établissement et analyse de listes de fréquences de termes dans l'ensemble du corpus étudié. Les points saillants révélés par la lecture des tableaux de fréquence sont ensuite interprétés dans leur contexte phraséologique, grâce à l'usage du concordancier.

La première étape consiste en un marquage des « mouvements » et en une analyse dans le cadre d'une stratégie rhétorique : ici, un exemple de résumé de thèse en sciences des matériaux, rédigé par un locuteur natif en anglais :

Frictional Resistance of Concrete on a Geocomposite Material,

[Cadre] The frictional resistance provided by the underlying layers of a concrete slabs on ground has a large influence on the cracking behavior. [Cadre] [Prop] To adequately judge the quality of performance of the slabs, it is imperative to investigate the frictional resistance to movement and other relations of the concrete and sub-layer system. It's also crucial to understand exactly how to introduce a drainage layer to the slab without adversely affecting the cracking pattern of the concrete by dramatically altering the resistance provided by the underlying layers. [Prop]. [Méthode] Frictional resistance of sub-grade

systems which incorporate the geocomposite material Tendrain were tested in a laboratory setting. The testing scenario included a predetermined number of blocks which were cast on top of Tendrain and then tested at varying ages of the concrete. Some additional tests were also conducted to determine any changes in the concrete shrinkage or compressive strength due to water seepage before setting of concrete. [Méthode]. [Résultats] Results from this study show: Tendrain has no significant affect on the compressive strength of the concrete, it some what reduces the amount of concrete shrinkage due to seepage and its frictional behavior is comparable to other sublayer systems. The connections between concrete and Tendrain found in this study are the preliminary steps to formulating slab design which utilizes the material as a sublayer. This study, along with future subgrade reaction analysis, can be used to predict elastic deflections and stresses for concrete slabs on ground. [Résultats]

Quatre mouvements rhétoriques se succèdent : l'auteur pose d'abord le cadre théorique, puis formule sa proposition de recherche, expose la méthode suivie, et enfin annonce les résultats obtenus. On voit que ces mouvements sont marqués notamment par des changements de temps successifs (cf. verbes soulignés).

Après le marquage des mouvements rhétoriques dans chacun des textes, on établit, à partir du corpus de résumés, des tableaux de fréquence comparatifs entre L1 et L2 et entre disciplines, avec un tri par catégorie syntaxique, sans lemmatisation.

2.3.1 Premier exemple d'extraction de tableaux de fréquence

Comparaison des occurrences de verbes étudiées grâce à des extractions de fréquence avec ANTCONC, à partir du corpus « Sciences des matériaux » en L1 et L2.

| Rank L2 | Occs. L2 | Form L2 | Occs. L1 |
|---------|----------|-----------------|----------|
| 1 | 63 | <i>is</i> | 93 |
| 2 | 29 | <i>are</i> | 53 |
| 3 | 21 | <i>have</i> | 0 |
| 4 | 19 | <i>used</i> | 19 |
| 5 | 17 | <i>been</i> | 5 |
| 6 | 16 | <i>loading</i> | 12 |
| 7 | 14 | <i>has</i> | 8 |
| 8 | 13 | <i>be</i> | 21 |
| 9 | 10 | <i>observed</i> | 1 |
| 10 | 10 | <i>studied</i> | 1 |
| 11 | 10 | <i>was</i> | 9 |
| 12 | 9 | <i>can</i> | 15 |
| 13 | 7 | <i>made</i> | 3 |

| | | | |
|----|---|-----------------|----|
| 14 | 7 | <i>opening</i> | 2 |
| 15 | 7 | <i>propose</i> | 1 |
| 16 | 7 | <i>proposed</i> | 8 |
| 17 | 7 | <i>repairs</i> | 0 |
| 18 | 7 | <i>were</i> | 6 |
| 19 | 6 | <i>allows</i> | 5 |
| 20 | 6 | <i>based</i> | 12 |
| 21 | 6 | <i>bending</i> | 1 |
| 22 | 6 | <i>carried</i> | 0 |
| 23 | 6 | <i>show</i> | 4 |
| 24 | 6 | <i>using</i> | 6 |

Tableau 1: Repérage des écarts les plus marquants entre les occurrences de verbes

L'analyse de l'écart de fréquence entre le nombre d'occurrences de « *have* » en L1 et en L2 grâce à la fonction « concordances » de ANTCONC au niveau de la phrase fait apparaître que les 21 occurrences de « *have* » en L2 incluent 9 occurrences de « *we have* » et 18 usages du *present perfect*.

L'étude de ces occurrences de « *we have* » dans l'ensemble des textes de L2 fait apparaître que les auteurs francophones recourent massivement au *present perfect* dans l'ensemble de leurs résumés, quel que soit le mouvement rhétorique exprimé. Ceci contredit l'usage constaté en langue 1, où les locuteurs natifs de l'anglais utilisent le *preterit* pour décrire leur méthode, dans une logique narrative.

Cette différence peut être interprétée comme un effet de calque de l'usage du passé composé en français, dont on sait qu'il n'a pas la même fonction que le *present perfect* anglais. Cette particularité de l'expression des locuteurs de langue 2 a pour effet involontaire de brouiller les limites des mouvements rhétoriques, dont nous avons vu qu'elles étaient notamment marquées par des changements de temps.

2.3.2 Deuxième exemple d'analyse réalisée à partir de comparaison de tableaux de fréquence d'occurrences de termes

Le tableau de fréquence d'occurrences de termes, comparant les corpus réalisés en L1 et L2 en sciences des matériaux, fait apparaître un écart important entre le nombre d'occurrences de « *analysis* » en L1 (30 occurrences) et L2 (4 occurrences).

L'étude du fonctionnement de ce terme dans son contexte phraséologique, grâce au concordancier, permet le constat suivant :

En langue 1, le terme « *analysis* » occupe deux fonctions essentielles : soit il est « tête » de collocation spécialisée, soit il représente une opération mentale liée à la recherche.

Dans les exemples suivants :

[9] *a geometrically nonlinear shell finite element analysis procedure*

[10] *a-finite element shell analysis*

la collocation relève de la terminologie de spécialité ; incluse dans le mouvement « Méthode », en position de complément, occupant la fonction de Rhème (Halliday).

En revanche, dans cet exemple :

[11] *Discrete modeling of crack propagation deals with the analysis of fracture processes in which the geometries of individual cracks are considered explicitly as integral parts of the analysis.*

le terme désigne un processus mental lié à la recherche, inclus dans le mouvement « Proposition » ou « Résultats », le plus souvent en position sujet, occupant la fonction de Thème.

Le fonctionnement du terme dans les rares occurrences de « *analysis* » en L2 est relativement similaire mais présente néanmoins des différences notables dans son rôle en contexte.

Dans les exemples suivants :

[12] *powder X Ray diffraction studies and EDX analysis*

[13] *electron diffraction in a TEM, infrared spectroscopy and thermogravimetric analysis*

l'efficacité de la collocation spécialisée est compromise par l'usage de la conjonction « et » dont l'interprétation est ambiguë.

Enfin, en ce qui concerne ces exemples :

[14] *Moreover, the analysis of their characteristics and their comparison with similar phosphates highlighted the relationships existing...*

[15] *We propose a total modelling of the delayed strains, based on the phenomenological analysis of the strains under load according to the shrinkage*

Dans le premier cas, le terme « *analysis* » est employé dans le mouvement « Proposition », dans le deuxième dans les Résultats : tous deux occupent une position de complément et une fonction de Rhème.

La comparaison des usages et contextes collocationnels du terme en L1 et L2 amène aux conclusions suivantes :

En L1 : « *analysis* » joue efficacement un rôle de « cristallisation » d'une collocation spécialisée ou désigne une opération mentale objet de la recherche.

En L2 : « *analysis* » joue les mêmes rôles, mais la cristallisation de la collocation est incomplète et donc moins efficace ; lorsqu'elle désigne une opération mentale de recherche, l'analyse n'est plus l'objet de la recherche, mais son instrument.

2.4 Des stratégies de communication différentes

L'ensemble des données relevées et analysées permet d'établir des différences importantes concernant non seulement l'usage des temps, la valeur sémantique des termes de la langue scientifique générale, mais aussi le choix lexical des verbes et des adjectifs et adverbes modalisants, différences entre les réalisations de textes en langue 1 et langue 2, mais aussi entre disciplines.

Ces écarts doivent être interprétés dans le cadre d'une stratégie de communication, en fonction de la répartition de ces marqueurs dans les « mouvements » rhétoriques. Elles doivent être évaluées en regard d'une efficacité communicationnelle fondée sur la capacité de l'auteur à créer chez son lecteur :

- a) un effet de cohérence ;
- b) une impression de légitimité ;
- c) la reconnaissance de la construction d'une « autorité ».

Il doit être bien clair qu'il ne s'agit pas d'établir une liste d'erreurs commises par les locuteurs non natifs, en faisant des productions des locuteurs natifs une norme à suivre.

Au contraire, il s'agit de savoir identifier des stratégies différentes au service des mêmes objectifs et d'en évaluer la réussite relative.

Ainsi, s'il est indéniable que l'usage du *present perfect* en L2 tout au long du résumé nuit à l'effet de cohérence, le choix du pronom *we* en L2, préféré à la domination du passif en L1, relève d'un choix de construction d'autorité différent mais non moins efficace.

De même, si une construction inefficace de collocations autour du terme « *analysis* » en L2, compromet l'effet de légitimation construit par une succession de collocations reconnues par les spécialistes du domaine, le fait de désigner l'opération d'analyse comme instrument plutôt que comme objet de recherche en L2 peut dénoter une vision différente de la recherche, non moins légitime.

Linguistique de corpus et analyse de discours se complètent, permettant finalement d'établir une typologie représentative des différentes stratégies de communication adoptées par les auteurs d'*abstracts* de thèse dans une même discipline, et en fonction de leur origine linguistique.

Les tableaux de fréquence fonctionnent comme des sortes de « tableaux de bord » où les écarts identifiés servent de révélateur des différences de fonctionnement lexical et sémantique. Ces différences ne peuvent être interprétées que grâce à une étude en contexte phraséologique et en les replaçant dans le cadre de l'analyse rhétorique faite pour chaque texte.

Ce type d'analyse permet de combiner l'objectivité des données obtenues sur l'ensemble du corpus et la finesse de l'analyse de discours.

Du point de vue de ses applications didactiques, cette approche permet de proposer la construction d'outils d'évaluation fondés sur une compréhension profonde du fonctionnement d'un genre et non sur une conception normative.

La compréhension des règles de fonctionnement d'un genre permet de montrer que la traduction de sa propre langue vers une langue étrangère ne consiste pas à transférer du texte, mais à construire un objet de discours équivalent dans sa fonction spécifique de communication.

3 Conclusion : conscience linguistique et choix traductologiques

On voit donc bien que le développement d'une conscience linguistique critique chez le traducteur s'avère indispensable afin de recréer dans la langue cible l'intention de l'auteur et sa réception par le lecteur. Rappelons en outre que les traducteurs ont souvent à traduire des textes qui ne respectent pas nécessairement les conventions de la langue source. La linguistique de corpus, en ce qu'elle ouvre une perspective large sur la diversité des réalisations de la langue, contribue à la CRÉATION de la conscience linguistique nécessaire pour affronter ces situations.

Nous pouvons donc conclure à l'efficacité sur le plan de la recherche comme sur le plan des applications de la combinaison des approches traductologiques et de linguistique de corpus.

The higher level of achievement is a contribution to the evaluation of the text
(Halliday 1985 : XV)

Bibliographie

- Aston (G.), 1999 : « Corpus use and learning to translate », dans *Textus*, 12, p. 289-313.
Baker (M.) 1999 : « The role of corpora in investigating the linguistic behaviour of translators », dans *International Journal of Corpus Linguistics*, n° 4, p. 281-298.

- Beeby (A.), Rodriguez (I.), Sánchez-Gijón (P.), éd., 2009 : *Corpus Use and Translating*, Amsterdam, Benjamins.
- Bernardini (S.), Castagnoli (S.), 2008 : « Corpora for translator education and translation practice », dans Yuste Rodrigo (E.), éd., *Topics in Language Resources for Translation and Localisation*, Amsterdam, Benjamins.
- Biber (D.), 2006 : « University language : a corpus-based study of spoken and written registers », dans *Studies in corpus linguistics*, ISSN 1388-0373 ; vol. 23. Amsterdam, J. Benjamins.
- Biber (D.), Connor (U.), Upton (T.) : 2008, *Discourse on the move*, Amsterdam, J. Benjamins.
- Bowker (L.), Pearson (J.), 2002 : *Working with Specialised Language: a Guide to Using Corpora*, London, Routledge.
- Gledhill (Chr.), 2000 : *Collocations in science writing*, Gunter Narr Verlag Tubingen.
- Halliday (M.A.K), Martin (J.R.), 1993 : *Writing science: literacy and discursive power*, University of Pittsburgh Press.
- Halliday (M.A.K.), 1994 : *An Introduction to Functional Grammar*. E. Arnold, 2^e éd. (1^{re} éd. 1985).
- Hickey (L.), 1998 : « Perlocutionary equivalence: marking, exegesis and recontextualisation », dans Hickey (L.), éd., *The Pragmatics of Translation*, Clevedon, Multilingual Matters.
- Hoey (M.), 1991 : *Patterns of lexis in text*, Oxford University Press.
- Hoey (M.), 2005 : *Lexical priming*, Routledge.
- Hunston (S.), Thompson (G.), 1999 : *Evaluation in text: authorial stance and the construction of discourse*, Oxford Linguistics.
- Hyland (K.), 2000 : *Disciplinary Discourse: Social Interactions in Academic Writing*, Harlow, Pearson Education Limited.
- Kübler (N.), 2003 : « Corpora and LSP translation », dans Zanettin (F.), Bernardini (S.), Stewart (D.), éd., *Corpora in Translator Education*, Manchester, St Jerome.
- Newmark (P.), 1988 : « Pragmatic translation and literalism », *TTR: traduction, terminologie, rédaction*, 1/2, p. 133-145.
- Stubbs (M.), 1996 : *Text and Corpus Analysis*, Oxford, Blackwell.
- Swales (J.), 1990 : *Genre Analysis: English in academic and research settings*, Cambridge University Press.
- Wable (T.), Holzem (M.), 2004 : « Structure textuelle et Communication scientifique », 11 mai 2004, http://halshs.archives-ouvertes.fr/sic_00001333/.
- Zanettin (F.), Bernardini (S.), Stewart (D.), éd., 2003 : *Corpora in Translator Education*, Manchester, St-Jerome.

Rosa Maria Fréjaville,
Département d'espagnol, portugais et catalan,
Université Jean Monnet,
Saint-Étienne,
France,
rosa-maria.frejaville@wanadoo.fr.

Le corpus en analyse terminologique de domaines composites et interdisciplinaires

Résumé

Dans cette communication, nous abordons la question de la constitution de corpus textuels à la lumière de la théorie communicative de la terminologie. Nous soulignons l'importance de la relation entre la communication en contexte de spécialité et les critères de sélection, d'organisation et de structuration d'un corpus textuel bilingue d'un domaine composite et interdisciplinaire pour la réalisation d'un programme terminologique. Pour ce faire, nous décrivons la méthodologie utilisée dans le traitement du domaine de la pollution, sa structure au sein d'une théorie, son circuit de communication en contexte spécialisé ainsi que sa mise en corpus, étape nécessaire pour son traitement terminologique.

Mots-clés: corpus textuel; communication spécialisée; domaine; discours.

Introduction

La méthode du corpus en analyse terminologique nous conduit inévitablement à repenser la méthode de traitement terminologique des domaines d'expérience longtemps utilisée par la théorie générale de la terminologie (TGT). Or, le traitement terminologique doit tenir compte non seulement du système de communication spécialisée mais également des textes, des discours et des termes qui définissent la « carte » de tout domaine. La linguistique de corpus apparaît ainsi comme l'interlocuteur incontournable de la théorie communicative de la terminologie (TCT) qui répond, le mieux, aux besoins d'analyse de domaines complexes tels que la pollution, domaine d'expérience relativement récent et très représentatif de la modernité technologique.

La TCT donne un relief particulier à la notion d'interdisciplinarité, qui est le fruit des contributions des théories du langage, de la connaissance, de l'information et de la communication. La TCT valorise donc les dimensions cognitive, linguistique et communicative qui structurent un domaine par une approche multidimensionnelle et « variable ».

1 Le domaine d'expérience

Un domaine d'expérience, ou selon Gaudin « secteur d'activité en fonctionnement », a un comportement sociodiscursif, il vit en discours, il est soumis à la variation et à la variabilité. Dans ce sens, son langage spécialisé ou « langage en contexte de spécialité » comme l'a remarqué, très judicieusement, Rute Costa lors de ces journées LTT de Lisbonne, se définit par un univers discursif, un circuit d'information et de communication et un système matriciel de textes.

Un domaine est le produit d'un programme de recherche, mis en scène pour définir une théorie et l'ériger en paradigme. Dans ce sens, la spécialisation de tout programme de recherche entraîne des « connaissances réflexives » : production de connaissances, décontextualisation, recontextualisation par le mouvement interne des concepts allant de la préparation/incubation à la vérification par un circuit très complexe de communication. L'organisation des savoirs est dépendante également des modèles et des contextes socioculturels sur lesquels elle s'appuie. Ce sont bel et bien les éléments de connaissance qui sont à l'origine du renouvellement et de la création des savoirs. Au départ, c'est par le biais des concepts, éléments identifiables par des propriétés et des relations, que la connaissance se structure en savoirs. Cette structuration dépend, néanmoins, d'un processus d'avancement effectif et des mécanismes d'auto-évaluation et d'hétéro-évaluation. Ainsi, les connaissances, en raison même de leur caractère grégaire, social et universel, font partie des théories et justifient le besoin de paradigmes.

1.1 Le domaine d'expérience de la pollution

Dans une perspective diachronique, mettant obligatoirement en évidence des aspects d'interdisciplinarité, le domaine d'expérience de la pollution relève d'un système que nous pouvons désigner « double triangle à vases communicants » qui met en évidence l'ontogenèse constitutive des réseaux conceptuels de l'écologie, de la pollution, entre autres secteurs liés à l'étude du milieu et de l'environnement. Prenons le schéma en exemple (figure 1) :

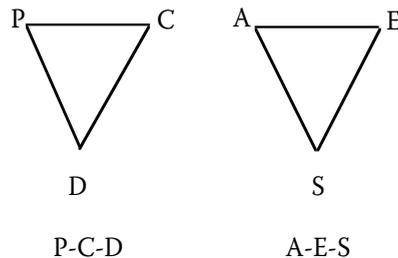


Figure 1

Dans le premier triangle, on trouve les concepts de « Producteur-Consommateur-Décomposeur » ; dans le deuxième triangle les concepts de « Air-Eau-Sol ». Il s'agit

du double triangle conceptuel en écologie composé de deux ensembles conceptuels cohérents en intrarelation et en interrelation et qui ont rendu possible le concept fondateur de l'écologie moderne, celui d'*écosystème*. Ce schéma rend perceptible la notion de concepts fédérateurs et la notion, par ailleurs, de nomadisme de concepts.

Lié à l'espace (triangle air-eau-sol), le *milieu* a été, d'abord, un objet immuable où circulaient les êtres vivants. Ensuite, il devient un ensemble de composants et finalement un groupe de facteurs qui expliquent la présence ou l'absence d'êtres vivants. Lié au vivant (triangle producteur-consommateur-décomposeur), le milieu a été, tout d'abord, qualifié d'harmonieux ensuite il a été défini comme un système de ressources, pour devenir, finalement, un élément abiotique et biotique, facteurs interagissant pour accomplir les besoins du vivant. Ceci s'est avéré fondamental pour l'interprétation du concept *écosystème* (ensemble biocénose et biotope). De là au terme *environnement* (spécialité au carrefour des sciences exactes, naturelles et humaines) il n'y a eu qu'un pas pour arriver à l'*altéralogie* en 1972.

De plus, les termes *écologie*, *environnement* et *pollution* sont des termes officiellement normalisés depuis les années 1970. L'AFNOR NF X30-001 les présentent comme suit :

- écologie *Ecology* – science traitant des relations des organismes vivants entre eux et avec leur milieu ;
- environnement *Environment* – ensemble, à un moment donné, des agents physiques, chimiques et biologiques et des facteurs sociaux susceptibles d'avoir un effet direct ou indirect, immédiat ou à terme, sur les organismes vivants et les activités humaines ;
- pollution *Pollution* – 1° introduction, directe ou indirecte, d'un polluant dans un milieu déterminé ; 2° résultat de cette action : présence de ce polluant dans ce milieu.

La plupart des scientifiques font toujours référence à la première définition de *pollution*, celle du Rapport du Conseil sur la qualité de l'environnement de la Maison Blanche (1965), comme étant la plus précise et la plus apte à de nouvelles adaptations conceptuelles. Le terme *pollution* y est défini comme étant : « une modification défavorable du milieu naturel qui apparaît en totalité ou en partie comme un sous-produit de l'action humaine, au travers des effets directs ou indirects altérant les critères de répartition des flux d'énergie, des niveaux de radiation, de la constitution physicochimique du milieu naturel et de l'abondance des espèces vivantes. Ces modifications peuvent affecter l'homme directement ou au travers des ressources agricoles, en eau et en produits biologiques. Elles peuvent aussi l'affecter en altérant les objets physiques qu'il possède ou les possibilités récréatives du milieu ». Le terme *pollution* a eu dès le départ le sème intrinsèque de l'altération, concept fondateur de l'*altéralogie* qui sera donc la partie de l'écologie consacrée à l'étude des altérations de l'environnement.

Si par l'analyse diachronique nous pouvons mettre en évidence les rapports entre l'écologie (domaine abstrait-concret) et la pollution, c'est par le corpus que nous pouvons présenter son système conceptuel, voire son univers « cognitif » dont sa thématique, son champ d'action et son univers de discours (l'emploi et l'usage, le système langagier et les variantes discursives).

Le domaine de la pollution est un domaine interdisciplinaire, multidisciplinaire et transversal. Interdisciplinaire parce qu'il est le produit d'une combinatoire sélective de concepts à partir d'autres champs de connaissances et qui l'a conduit à devenir un nouveau champ de recherche déterminé par un objet d'étude et une méthode qui lui sont propres. Multidisciplinaire parce que le domaine de la pollution subit l'intervention d'autres disciplines, notamment de la biochimie, sans que cela le détourne de sa spécificité. Il possède des traits de transversalité parce qu'il peut être analysé selon des perspectives différentes : scientifique, juridique, politique, médical, etc., constituant un objet d'étude transdisciplinaire.

Toute la difficulté pour son analyse réside à construire, au départ, avec fidélité, sa carte conceptuelle en établissant les relations entre ses concepts fondateurs. Travail accompli, le domaine de la pollution est un ensemble ontologique qui a trait, tout d'abord, à l'écologie et à l'environnement, et ensuite, par des degrés de spécialisation, à des secteurs plus ou moins techniques et technologiques. Ce qui ressort de cette articulation des théories ce sont les réseaux conceptuels entre les domaines de l'écologie, de l'environnement et de la pollution. Cette articulation atteste l'importance de l'incommensurabilité de rapports entre les programmes de recherche (Lakatos 2001), promoteurs de nouveaux paradigmes scientifiques, alimentés par des moyens éditologiques (figure 2) :

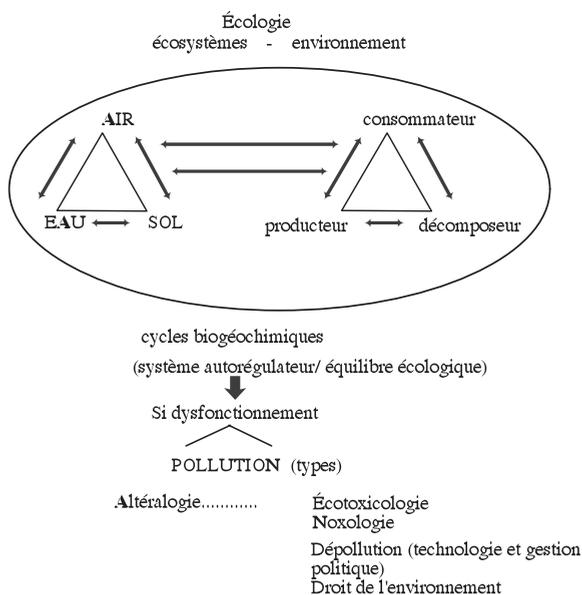


Figure 2

On remarque aisément que la pollution est un domaine d'expérience à forte coordination entre contributions scientifiques, applications technologiques et productions industrielles (un système de type R & D).

Pour arriver à ces résultats, il faut que l'analyse historique et la démarche de catégorisation puissent contribuer à décrire l'espace conceptuel du secteur de la pollution rendant visibles ses structures cognitives, organisées en plusieurs ensembles/termes interdépendants.

Les données de connaissance de la pollution peuvent, de la sorte, aisément être distribuées en catégories et réparties en classes, englobant l'ensemble de ses propriétés communes et complémentaires fort utiles pour le repérage, le classement des termes et la libération du sens, tout en permettant le repérage des relations qu'elles entretiennent. On remarque, également, que cette perspective thématique englobe les deux langues choisies, le portugais et le français.

Ainsi, pour ces deux univers langagiers, avec des concordances au niveau terminologique et des variations plus ou moins nettes au niveau discursif, on relève six grands sous-ensembles : *pollution atmosphérique* ou *pollution de l'air*, *pollution de l'eau*, *pollution des sols*, *pollution sonore*, *pollution par l'odeur*, *pollution industrielle*; des classificateurs tels que les *déchets* et les *polluants* ainsi qu'un sous-ensemble opposant déséquilibre écologique et équilibre écologique dont : l'équilibre écologique par le droit, la *dépollution*, l'*antipollution* et les réseaux *contamination*, *bioaccumulation*, *concentration* et *catastrophes écologiques* qui démontrent les interrelations sémantiques des deux triangles exposés en attestant son caractère scientifique, économique, juridique et technologique.

On s'aperçoit donc que pour bien connaître un domaine, issu et soumis à un programme de recherche, il faut comprendre son univers de discours (dont les terminologies), les interactions de ses attributs linguistiques et langagiers, caractéristiques de son comportement scriptural et communicatif. Le discours de la pollution est une structure descriptive et constructive, détectable à partir de thèmes organisés de façon cohérente, permanente et rigoureuse. Son discours est le résultat, somme toute, de la conjonction d'une structuration spécifique de concepts, et d'un certain type d'énonciation.

2 Ordre scriptural, communication spécialisée et corpus textuel

Le domaine étant ainsi identifié, il faut attester la pertinence des choix effectués par l'identification de son circuit d'information et de communication. Ceci nous amène directement à la question de la communication spécialisée qui est le deuxième maillon nécessaire au travail terminologique.

La communication spécialisée n'a pas cessé de prendre de l'ampleur grâce, notamment, aux apports de la théorie de l'information et de la communication : circulation

plus rapide des idées et des connaissances et de surcroît une plus grande rapidité dans l'évolution des spécialités qui aboutit à des conséquences entre le savoir partagé et le savoir réparti. Véritable apport épistémologique pour l'analyse terminologique, la communication spécialisée est le préalable à la construction de corpus textuels répondant à des besoins spécifiques qui, en pollution, se définit par un vaste programme éditologique.

Que veut-on dire par communiquer en langage spécialisé ? C'est tout simplement transmettre les résultats d'un programme de recherche, l'intégrer dans un programme d'enseignement, le diffuser auprès du grand public. Il s'agit en définitive de construire des lieux de débat et de promouvoir les transferts d'informations qui prennent forme par la communication informelle, puis s'instrumentalisent par la communication institutionnelle et deviennent des connaissances par la communication publique, dans différents degrés de vulgarisation.

La logique d'investigation et la logique de l'exposition sont interactives, le système de communication spécialisée leur obéit dans la mesure où un chercheur et/un programme de recherche, pour exister, doivent faire l'objet de publications qui, à leur tour, donnent existence officielle aux domaines, voire aux disciplines et à leurs chercheurs. En vérité, la communication spécialisée est « une circulation textuelle intense et multiforme » (Gaudin 1933 : 33).

Le support écrit est sûrement le moyen le plus utilisé par la communication spécialisée et l'outil privilégié de la transmission des connaissances, englobant tous les textes édités en tant que systèmes d'unités communicationnelles qui sont autant de discours.

La communication spécialisée est donc véhiculée par les moyens scripturaux qui sont des réalisations graphématiques du texte de spécialité (en opposition aux réalisations phonématiques) et à travers lesquels on peut tenir compte des questions ayant trait à l'omniprésence des reformulations, de l'altération (états d'équivalence) et d'altérité (états de différence).

L'ordre scriptural (Jacobi 1984) est représenté par une littérature spécialisée qui est restreinte et publique, informative et formative, constituante et non constituante. C'est par les différents types de textes que la science peut parler, en quelque sorte, d'un même sujet à travers des structures textuelles et des registres différents. Ainsi, la *scripturation*, processus de mise à l'écrit par les reformulations permettant la transmission, diffusion et dissémination des savoirs, permet au terminologue de connaître les schémas argumentatifs, et par là même les mouvements des concepts et des termes d'un domaine. Le terminologue peut également mieux cerner, dans un état précis de son évolution, la structurelle conceptuelle et terminologique du domaine traité. Jacobi, d'ailleurs, qualifie d'essentielle la dimension *scriptovisuelle* de l'écrit scientifique en mettant en avant l'importance de la documentation pour tout travail terminologique.

La communication spécialisée étant bien évidemment un système qui résulte de l'application des terminologies des domaines, présente des caractéristiques particulières : elle est impliquée dans le processus d'interaction communicative, dans la thématique traitée, dans les relations entre interlocuteurs soumis aux contraintes de l'espace et du temps ; d'où découlent des spécificités relatives à la terminologie et aux types de textes produits. C'est pourquoi la notion même de corpus textuel doit être indissociable du processus de communication spécialisée.

2.1 Textes et communication spécialisée de domaines composites et interdisciplinaires

Les textes de spécialité sont les supports privilégiés de l'analyse terminologique. Unités fondamentales de la communication spécialisée, ils facilitent la réalisation d'une analyse terminologique à la fois globale et approfondie de domaines composites. Par nature, unités de fonctionnement d'un système terminologique, ces textes peuvent, par conséquent, être analysés en fonction de leur nature, leurs spécificités et leurs finalités pragmatiques.

C'est la thématique qui fait le départ entre les textes de spécialité et les autres, permettant non seulement d'établir les différentes disciplines conditionnées par la forme, la structure et le style des textes mais aussi des approches de type contrastif et de type méthodologique. En effet, c'est le thème qui permet d'identifier un domaine d'expérience par rapport à un autre. Et selon les différents secteurs ou degrés de variation d'un même domaine, on pourra dégager plusieurs approches descriptives. La terminologie utilisée, dépendante de sa thématique, peut être plus ou moins opaque et révéler plusieurs degrés de variation, selon ses finalités communicationnelles. De plus, chaque matrice de textes est déterminée également par des spécificités propres à la langue véhiculée et qui sont, bien évidemment, terminologiques et rhétoriques. Ces spécificités sont terminologiques et rhétoriques, caractérisées par des traits argumentatifs, phraséologiques et sémantico-syntaxiques. Ainsi, le plan pragmatique concerne le degré de compréhension interlinguistique entre les différents interlocuteurs d'une même langue ou de langues différentes. Ces textes utilisent des ressources terminologiques et discursives particulières, parfaitement adaptables à la diversité des situations de communication et à la singularité des compétences. Les textes de spécialité sont caractérisés par une sémantique globale qui les identifie comme étant des textes hautement référentiels. Dans ce sens, on peut remarquer : les spécificités formelles qui leur confèrent divers degrés de restrictions structurelles ; les caractéristiques spécifiques des terminologies utilisées ; les thématiques souvent utilisées à travers des systèmes alternatifs intégrés du genre reformulations, paraphrases, citations, modèles de traits linguistiques, entre autres. Or, bien sûr, les textes de spécialité ont recours à des procédés de type matriciel. Nous sommes, donc, en présence de textes très structurés, en adéquation avec le thème, le rhème et les destinataires. C'est pourquoi il faut tenir compte des éléments qui leur accordent un statut spécialisé : les différentes relations entre système conceptuel et structure terminologique, entre discours et destinataires, entre les sources documentaires et le référentiel.

Si on se tient, d'une part, aux différentes étapes du circuit de la communication spécialisée, et, d'autre part, en considérant que le texte de spécialité est un maillon dans le plurisystème graphique et qu'il est le produit premier d'emboîtements reformulateurs, divers modèles peuvent être dénombrés.

En amont du circuit de communication spécialisée, on trouve des textes du genre notes de laboratoire de très difficile accès au terminologie, car il s'agit souvent d'un début de recherche ou de travaux en cours; des exposés, des synthèses scientifiques; des rapports scientifiques et techniques qui sont des textes hautement référentiels, caractérisés par une grande densité terminologique qui s'adressent à la communication entre pairs, normalement du même domaine ou du même secteur d'expérience. Les travaux achevés, ces textes sont mis à la disposition d'un large public qui dans un premier temps est composé de spécialistes du même domaine, ensuite des spécialistes de domaines différents, puis d'un public en formation, et finalement du grand public.

En d'autres termes, cette circulation de textes, intense et organisée a des finalités scientifiques, pédagogiques et vulgarisatrices. Scientifiques parce que les textes diffusés vont contribuer au débat scientifique et, donc, à l'affirmation et à la reconnaissance du domaine; pédagogiques parce que ces textes contribuent à la création et à la dissémination des disciplines; et vulgarisatrices parce que ces textes servent à la diffusion de la connaissance et, de surcroît, à la diffusion des terminologies dans l'univers du quotidien.

Pour ce qui est du domaine de la pollution, et par ordre décroissant de spécialisation, on peut dégager plusieurs modèles: les articles créatifs ou des articles scientifiques qui sont soumis à certaines fonctions prépondérantes telles que les fonctions sociales de diffusion et de promotion de connaissances; les thèses, les revues spécialisées, les manuels, les livres de vulgarisation, les documents officiels et gouvernementaux, les dépôts de brevets, les études, les programmes; allant jusqu'à un foisonnement de productions éditoriales telles que les bulletins, les lettres, les guides, les programmes technologiques, les catalogues, les brochures et les dépliants techniques ou à visée sociale.

Ces quelques modèles textuels témoignent de la nature ouverte du circuit communicatif de la connaissance et par là même de l'impossibilité de concevoir les spécialités en tant que domaines cloisonnés. La pollution, ou toute autre thématique, peut être considérée sous des angles divers et/ou complémentaires en fonction des types de discours et des divers styles discursifs (ou niveaux de spécialisation). En fin de compte, on est en présence de variables déterminées par l'adéquation de la thématique à la fois aux finalités et aux différents utilisateurs. Dans un cadre strictement scientifique, le discours fait la description de l'objet qui est le fondement de la recherche et du travail du spécialiste. Dans un cadre plus technologique, il s'agit d'actions où c'est la rentabilité qui compte. Tout ceci revient à dire qu'une grande importance est accordée au type de discours axé sur la sémantique du « faire »: faire savoir, savoir-faire et faire faire.

Le système textuel du domaine d'expérience de la pollution ne s'écarte donc pas de la règle de construction des discours spécialisés, car il est fait de descriptions, d'énumérations, de calculs, de classifications, d'argumentations, de prescriptions, de citations, de références. Il s'agit d'un ensemble de reformulations et d'une multitude de relations d'équivalence à divers degrés discursifs qui sous-tendent une variabilité de degrés de spécialisation. On est, en somme, en présence de différents contextes énonciatifs.

In fine, ces textes témoignent de l'état scriptural du domaine d'expérience de la pollution et peuvent, donc, alimenter son corpus: une structure de texte constituant du type scientifique et technologique et une structure de type économique, juridique, voire politique.

2.2 Le corpus textuel : support et apport du programme terminologique

Un corpus textuel doit donc rendre visible la nature et l'ossature d'un domaine: ses dimensions discursives, son contenu et sa forme, qui doivent être en adéquation avec l'évolution de son système conceptuel et l'élargissement de son champ d'investigation. De surcroît, le corpus peut être représentatif des différents degrés de spécialisation d'un domaine et, par conséquent, apporter d'indubitables informations sur son paradigme scientifique et, à l'évidence, sur son caractère composite et/ou interdisciplinaire.

Le corpus textuel est non seulement le support nécessaire à tout programme terminologique mais également un outil à l'encodage et aux performances du travail terminologique. Les textes de spécialité peuvent ainsi être identifiés par une double focalisation: une « vision externe » parce qu'ils donnent une vision de l'univers où les discours se sont produits; une « vision avec » parce qu'ils rendent visibles les discours produits dans cet univers.

Si les textes sélectionnés et répertoriés doivent être au départ le reflet de la *carte conceptuelle* du domaine, le corpus organisé doit pouvoir répondre à la problématique posée par le projet terminologique. Les fonctions d'un corpus constitué sont donc multiples: l'analyse de discours de spécialité, le traitement de champs conceptuels et de champs terminologiques, l'extraction et le traitement de *terminoclosures*.

Dans le programme terminologique d'un domaine composite tel que la pollution, le corpus textuel doit idéalement pouvoir décrire ses discours et sa terminologie en diachronie et en synchronie. En d'autres termes, pour satisfaire les étapes mêmes du travail terminologique, et tenant compte des différentes tâches qui lui sont imparties, une méthode de répartition d'un corpus global ou primaire doit être envisagée.

Prenant l'exemple de la pollution, la sélection du corpus global est la deuxième étape du travail terminologique. C'est-à-dire, après l'établissement du réseau de concepts et l'analyse des caractéristiques de son circuit de communication spécialisée, la

recherche de la documentation pertinente et en adéquation avec les résultats obtenus, peut être faite.

Une double approche onomasiologique et sémasiologique s'avère, donc, nécessaire, pour accomplir les étapes de la constitution du corpus textuel, qui dans le cas ici décrit concerne un corpus bilingue portugais-français de la pollution. Ainsi, les étapes pour la constitution du corpus textuel sont les suivantes :

- un inventaire exhaustif de la documentation à travers une analyse préalable des systèmes de classification sur la pollution (CDU et/ou thésaurus) pour la recherche de terminologie disponible suivi de la collecte de bibliographie par thèmes-matière et/ou réseaux de concepts; ainsi que la collecte de documents sur la pollution et l'environnement après contact avec les professionnels, les associations et les organismes (ADEME, AFNOR, agences de l'eau, etc.) crédibles et représentatifs;
- une analyse comparative de la documentation suivie d'une sélection d'ouvrages et d'une analyse de la pertinence et de la qualité des sources (dans les deux langues) avec validation par les spécialistes afin de tester l'homogénéité du corpus et dessiner le champ conceptuel de la pollution (dont les résultats attestent l'importance de la pollution en tant que champ d'expérience et champ d'activité);
- une analyse comparative des données répertoriées dans les textes avec les systèmes de classification suivie de l'établissement d'un premier schéma classificatoire du domaine d'activité à traiter, puis une comparaison des listes de termes dans les deux langues (portugais et français) avec création d'une liste d'exclusion de termes afin de préserver l'homogénéité du corpus bilingue.

Dans cette phase du travail terminologique, le grand ensemble de données recueillies est ensuite distribué en plusieurs corpus: un corpus textuel de référence global répondant au programme terminologique envisagé et dont la finalité est l'acquisition de la compétence cognitive du domaine; et un corpus textuel d'analyse, parallèle, servant au repérage, dépouillement et traitement terminographique (*cf.* Annexe 1). Cela rend plus aisée l'élaboration des tableaux classificatoires du domaine organisés en nomenclature et par catégories rendant visible son système conceptuel et terminologique ainsi que le traitement terminologique du domaine, c'est-à-dire, l'extraction et l'analyse des éléments terminologiques tels que les termes et les phraséologies, les contextes, les collocations, les reformulations, etc.

Conclusion

Un programme de recherche terminologique de domaines composites et interdisciplinaires, mettant en avant une méthodologie axée sur une étude préalable de son statut ontologique, voire ontogénique, rend plus aisé tout travail terminologique aussi bien au niveau de la sélection du corpus qu'au niveau de son traitement. De plus, tout travail terminologique, à l'heure actuelle, où la société dépend

fondamentalement d'une organisation consensuelle de l'information et de la communication, exige l'appui inconditionnel de corpus textuels.

En effet, la méthode du corpus, mettant en avant les textes de spécialité, et parce que l'analyse terminologique ne peut pas dissocier terme/concept et référent, représente un enjeu considérable en sémantique textuelle et en traitement automatique et/ou semi-automatique des langages spécialisés.

Damon Mayaffre (2005 : 5) met en évidence deux perspectives fortes qui voient dans le corpus, l'une *l'observatoire* d'une théorie *a priori* et l'autre un *observé dynamique*. Nous croyons également que le corpus textuel est un élément incontournable, car il est à la fois l'observatoire et l'observé nécessaires à l'accomplissement d'un programme terminologique concernant des domaines vastes, variés et complexes.

Nous souhaitons que ces quelques propos sur l'importance d'une méthodologie mixte, qui tient compte des spécificités de la communication spécialisée et de la terminologie sur corpus, contribuent à la réflexion sur le rôle et la place des corpus dans les programmes de recherche en terminologie et éclairent les enjeux de la diversité.

Bibliographie

- Baudet (J. Cl.), 1991 : « Éditologie et sociolinguistique », dans *Cahiers de Linguistique Sociale*, n° 18, p. 81-99.
- Brunet (E.), 1999 : *Logiciel Hyperbase, hypertexte statistique pour grands corpus: nouvelle version Mac et Windows sur cédérom*, Nice/Paris, InaLF/Champion.
- Cabré (M.T.), 1998 : *La terminologie – Théorie, méthodologie et applications*, Canada, PUO/Armand Colin.
- Cabré (M.T.), 1999 : *La Terminología – representación y comunicación*, Barcelona, IULA.
- Camlong (A.), 1991 : *Stablex pratique*, Toulouse, Teknea.
- Dupupet (M.) et Fréjaville (R.M.), 1997 : « Contribuições para a constituição da(s) terminologias(s) da ecologia », dans *Revista Brasileira de Linguística*, vol. 9, n° 1, p. 141-163.
- Fréjaville (R.M.), 2005 : *Étude terminologique du discours spécialisé de la pollution – vers la structuration d'un modèle de dictionnaire terminologique bilingue français-portugais*, Lille, ANRT.
- Fréjaville (R.M.), 2006 : « Ciência e terminologias: o domínio de experiência da Poluição », dans *Revista Brasileira de Linguística*, vol. 12, n° 1, p. 193-215.
- Gaudin (F.), 1993 : *Pour une socioterminologie, des problèmes sémantiques aux pratiques institutionnelles*, Rouen, PUR.
- Habert (B.) et al., 1997 : *Les linguistiques de corpus*, Paris, Armand Colin.
- Jacobi (D.), 1984 : « Du discours scientifique, de sa reformulation et de quelques usages sociaux de la science », dans *Langue Française*, n° 64, p. 38-52.
- Lakatos (I.), 2001 : *The methodology of scientific research programmes*, Cambridge, University Press.

Maingueneau (D.) et Cossutta (F.), 1995 : « L'analyse des discours constituants », dans *Langages*, n° 117, mars 1995, p. 112-125.

Mayaffre (D.), 2005 : « Rôle et place des corpus en linguistique : réflexions introductives », dans *Actes de journées d'études toulousaines*, p. 5-17.

Mortureux (M.F.), 1988 : « Vocabulaire scientifique et circulation du savoir », dans *Protée*, vol. 16, 3.

Annexe

1 Corpus documentaire de référence

| Textes | Nombre d'ouvrages |
|---|-------------------|
| Dictionnaires sur l'écologie, l'environnement et la pollution | 20 |
| Ouvrages sur l'écologie et l'environnement | 42 |
| Ouvrages sur la pollution et l'environnement | 34 |
| Programmes, études, rapports et manuels | 51 |
| Guides | 9 |
| Périodiques | 15 |
| Dépliants et brochures | 62 |

2 Corpus documentaires de base

- a. Corpus textuel informatisé (A) par traitement *STABLEX* pour l'analyse descriptive du discours de la pollution : 8 ouvrages en français et 7 ouvrages en portugais
- b. Corpus textuel informatisé (B) par traitement *Hyperbase* pour l'analyse du secteur de la pollution pour collecte des unités terminologiques ainsi que l'identification des relations sémantico-pragmatiques entre les termes : 34 ouvrages en français et 20 ouvrages en portugais.

*Lorraine Goeuriot,
Yukie Nakao,
Béatrice Daille,*
LINA CNRS UMR 6241,
Nantes,
France,
lorraine.goeuriot@univ-nantes.fr,
yukie.nakao@univ-nantes.fr,
beatrice.daille@univ-nantes.fr.

Analyse de la modalité dans un corpus spécialisé multilingue

Résumé

Avec l'expansion des données textuelles dans le domaine médical, des méthodes d'organisation sont nécessaires. Celles-ci reposent sur des informations qui reflètent le thème, le genre, le niveau de communication des documents. Dans cet article, nous cherchons à distinguer deux niveaux de communication du domaine médical : scientifique et vulgarisé. Nous effectuons alors une analyse automatique de la modalité, relation établie entre l'auteur et le contenu de son discours. Nous adaptons deux théories de la modalité définies sur la langue générale au domaine médical en français et japonais. Nous montrons que l'automatisation est possible, mais seule l'une des deux théories se montre vraiment efficace et pertinente.

Mots-clés : langues de spécialité ; modalité ; analyse automatique.

Introduction

Les quantités de données textuelles sont en constante augmentation. Les progrès technologiques et l'accès à Internet permettent maintenant à qui que ce soit de publier du texte sur le Web. Il est toutefois essentiel de disposer de méthodes pour explorer et organiser ces données. Les outils permettant de classer des documents à partir de collections ou du Web sont nécessaires afin d'exploiter les connaissances qu'elles véhiculent. L'une des tâches de la recherche d'information consiste à collecter des textes pertinents en utilisant des critères linguistiques : le projet PRINCIP (Vinot *et al.* 2003) a par exemple pour but de détecter des pages contenant des propos racistes.

La gestion documentaire et la classification automatique visent à organiser les informations de façon à procurer à un utilisateur un document pertinent, par rapport à son sujet et à son contenu. Dans le domaine médical, le niveau de connaissance de l'utilisateur est lui aussi très important pour rechercher un document, son lectorat doit donc être identifié : les patients cherchant des conseils

ou informations médicales ne devraient pas avoir accès à la littérature spécialisée, incompréhensible. Les portails médicaux, tels que CISMEF en France, classent les documents selon leur genre ou type, qui reflètent le « contexte de production ou d'usage » des documents (Habert *et al.* 2001).

Nous distinguons deux types de discours : scientifique et vulgarisé. Les documents scientifiques sont écrits par des spécialistes, à destination de spécialistes : l'auteur et le lecteur partagent le même niveau de connaissance. Les documents vulgarisés sont destinés au « grand public » et sont écrits par des spécialistes ou non-spécialistes, ils ont généralement une visée pédagogique. La classification de documents selon leur type de discours est utile afin de construire des lexiques de termes spécialisés et leurs équivalents vulgarisés (Elhabad et Sutaria 2007) ou pour améliorer la qualité de terminologies bilingues (Morin *et al.* 2007).

La relation établie entre l'auteur et le lecteur est particulièrement pertinente dans l'analyse d'un texte. Bien que les auteurs de textes spécialisés soient souvent des spécialistes du domaine, certains niveaux de communication peuvent être distingués, influençant le ton, le style ou le lexique utilisé par l'auteur. Dans le cas des documents scientifiques, la plupart des énoncés utilisent un style impersonnel, le style personnel apparaît uniquement lorsque l'auteur exprime son point de vue. Par exemple, dans la phrase

[1] Nous pensons que ce sont surtout les modalités de diagnostic de cancer du sein des femmes âgées qui sont en cause.

nous pensons exprime l'opinion de l'auteur concernant le diagnostic de cancer du sein. Quant aux documents vulgarisés, les deux styles sont utilisés. Par exemple, dans la phrase

[2] Vous pouvez craindre que d'autres membres de votre famille ne développent un cancer.

l'auteur s'adresse directement à son lecteur en supposant l'inquiétude de l'interlocuteur (patient) concernant la maladie. Dans ces exemples, nous soulignons la subjectivité de l'auteur et sa relation au lecteur, appelées « modalité ». L'analyse de la subjectivité peut en effet être pertinente afin de caractériser deux types de discours et nous allons nous intéresser à ces aspects de la modalité. Rares sont les travaux d'analyse des langues de spécialité s'intéressant à l'attitude du locuteur. Sionis (2002) étudie la modalité dans le discours scientifique dans le cadre de l'apprentissage de la langue de spécialité pour les locuteurs non natifs anglais. Pecman (2007) étudie la distribution de critères de la modalité dans le discours scientifique afin de créer une plateforme de gestion de ressources linguistiques scientifiques. Dans chacun de ces travaux, la notion de modalité est inspirée de travaux sur la langue générale.

Dans cet article, nous cherchons à déterminer si l'analyse de la subjectivité dans des documents issus du domaine médical peut permettre de distinguer les documents scientifiques des documents vulgarisés. Notre objectif est de démontrer qu'il existe

des marqueurs de modalité dans ces documents et que ceux-ci peuvent être détectés automatiquement.

De plus, nous nous intéressons ici à deux langues, le français et le japonais : la première appartient à la famille des langues indo-européennes tandis que la seconde, n'appartenant à aucune famille de langues existante, a son propre système langagier. Le multilinguisme pose problème dans le cadre de la fouille de texte et, dans notre étude, le choix des théories de la modalité doit être fait minutieusement pour être appliqué à deux langues. Néanmoins, la forte distance entre ces deux langues n'empêche pas des similarités linguistiques quant à l'expression de la subjectivité.

Nous présenterons dans la section 2 deux théories : la théorie locutive et la théorie *irrealis*. Notre étude s'est basée sur l'analyse d'un corpus comparable spécialisé en français et japonais, que nous présenterons en section 3. L'analyse de la modalité sur le corpus, la méthode et les résultats sont décrits en section 4. Le but de cette analyse est d'adapter les théories de la modalité aux deux langues ainsi qu'aux langages de spécialité. En nous basant sur une première analyse manuelle sur un échantillon du corpus, nous réalisons une automatisation de la reconnaissance des critères des deux théories de la modalité à l'échelle de notre corpus. Nous discuterons les résultats en section 6, en cherchant à déterminer si l'application de ces théories aux langues de spécialité en français et japonais est pertinente, mais aussi si ces théories sont effectivement automatisables.

1 La modalité

Un énoncé contient deux types d'information : une proposition et une attitude du locuteur par rapport à son propos. Cette attitude est appelée « modalité ». La modalité dévoile, entre autres, la subjectivité du locuteur et le lien avec son interlocuteur (Palmer 1986, Le Querler 1996). Dans l'énoncé

[3] Je crois qu'il pleut.

la proposition de l'énoncé est *il pleut*. La partie modale *je crois* indique l'attitude (croyance) du locuteur à son propre propos. Selon Palmer (1986), la typologie de la modalité permet d'identifier des catégories grammaticales entre différentes langues. Cette particularité est importante : nous nous intéressons à l'analyse d'un corpus comparable et bilingue. Dans cette section nous présentons deux théories de la modalité : la théorie de la modalité locutive et celle de la modalité *irrealis*.

Définie pour la langue française par Charaudeau (1992), la théorie locutive classe principalement les expressions selon des constructions verbales, contenant des expressions modales, en se basant sur leurs catégories sémantiques. Chaque catégorie est définie indépendamment de la grammaire française et nous supposons que l'application de cette théorie à d'autres langues permet de différencier les caractéristiques de chaque langue. Afin de compléter des détails des modalités en japonais, nous nous sommes basés sur les travaux de Miyazaki *et al.* (2002) et de

Masuoka *et al.* (1992). La théorie *irrealis*, présentée comme indépendante de la langue (Chung et Timberlake 1985, Givón 1994), a déjà été appliquée à plusieurs langues, parmi lesquelles le français (Sionis 2002) et le japonais (Tamachi 2007). Nous avons pris la classification de cette théorie proposée par Givón (1994). L'application de ces deux théories va de plus nous permettre d'évaluer s'il est possible de comparer deux langues sur une même base.

1.1 Modalité locutive

L'idée principale de la théorie de la modalité locutive de Charaudeau porte sur les actes locutifs classés en trois actes : l'acte alocutif, l'acte élocutif et l'acte délocutif. Avec l'acte allocutif, le locuteur implique son interlocuteur en lui imposant le contenu de la proposition

[4] Vous pouvez rentrer maintenant.

Cet acte est composé de neuf catégories telles que « injonction », « autorisation », « jugement », « suggestion ». Avec l'acte élocutif, le locuteur ne mentionne pas directement l'interlocuteur, mais lui-même

[5] Je pense qu'il fera beau demain.

Cet acte est composé de deux catégories telles que « savoir/ignorance », « opinion », « obligation ». Enfin, l'acte délocutif ne réfère ni au locuteur ni à l'interlocuteur

[6] Il est évident qu'elle n'était pas chez elle hier.

La théorie locutive nous permettant de déterminer la fonction des locuteurs dans l'énoncé, l'acte délocutif n'est pas traité dans notre travail.

Les modalités allocutives et élocutives sont généralement caractérisées par des verbes et des constructions verbales. Par exemple, la modalité de la suggestion se forme par des verbes tels que « suggérer », « conseiller » pour le français, « 勧める » (recommander) pour le japonais. La même modalité est également exprimée par la conjugaison de deuxième personne pluriel *-ons* pour le français et le suffixe verbal «-ましょう» (équivalent à *-ons* du français) pour le japonais.

1.2 Modalité *irrealis*

L'autre théorie de modalité, la modalité *irrealis* (Givón 1994, Nomura 2003), examine si la proposition dans un énoncé est réalisée ou non. Lorsque le locuteur croit que l'événement est réalisé, l'énoncé en question est catégorisé comme « Realis ». Lorsque le locuteur croit que l'événement n'est pas réalisé, l'énoncé est catégorisé comme « Irrealis ». La modalité *irrealis* proposée par Givón (1994) se caractérise pour la langue anglaise par des critères grammaticaux, par exemple les verbes au temps futur :

[7] Il connaîtra le résultat demain ;

les adverbes modaux :

[8] Je ne viens probablement pas ;

la commande :

[9] Vas-y ! ;

l'impératif de la première personne plurielle :

[10] Allons manger ;

les questions oui/non :

[11] Tu n'es pas venu hier, n'est-ce pas ?

et les auxiliaires modaux :

[12] Je dois retourner à la préfecture.

Afin d'adapter cette théorie à langue française, nous avons inclus le conditionnel et l'imparfait

[13] Si j'étais riche, j'achèterais une maison de campagne

dans la catégorie de la phrase adverbiale et le futur antérieur

[14] Elle aura manqué son train.

dans la catégorie du temps futur. En japonais, des formes composées sont des formes verbales idiomatiques

[15] « としい », « ce serait bien si... » équivalentes à la phrase adverbiale.

2 Le corpus d'étude

Afin d'étudier ces théories de la modalité, nous avons besoin d'un corpus multilingue spécialisé contenant des documents relevant de chacun des types de discours. Nous décrivons dans cette section notre corpus d'étude et sa constitution.

2.1 Présentation

Nous avons construit un corpus comparable dans les langues française et japonaise. Un corpus comparable est un ensemble de documents qui ne sont pas des traductions mais partagent un certain nombre de caractéristiques, telles que l'auteur, la période, le genre, etc. (Bowker et Pearson 2002).

Le choix de ces caractéristiques communes influe sur le degré de comparabilité du corpus : un haut degré de comparabilité garantit une homogénéité lexicale dans le corpus (Bowker et Pearson 2002), requis pour de nombreuses applications, telles

que l'extraction d'informations multilingues. Notre objectif est de construire un corpus représentatif d'un domaine de spécialité et des deux types de discours que nous étudions. Dans les corpus comparables spécialisés, les documents partagent généralement comme premier degré de comparabilité le domaine et le thème. Un thème commun permet de limiter le champ de documents pouvant être inclus et facilite ainsi la construction. Nous avons choisi le domaine médical et la thématique « diabète, alimentation et obésité ». Après collecte et analyse de documents, différents niveaux de communication ont été identifiés (Pearson 1998) : spécialiste-spécialiste, spécialiste-initié, semi-spécialiste-non-initié, enseignant-élève. À chacun de ces niveaux correspondent différentes caractéristiques linguistiques : selon le niveau de connaissance de l'auteur et du lecteur attendu, le style, le vocabulaire et la structure du document diffèrent. Malrieu et Rastier (2002) ont introduit différents niveaux de classification textuelle, correspondant chacun à un degré de granularité : discours, genre. En nous inspirant de ces niveaux de communication et de classification textuelle, nous avons limité notre analyse aux deux catégories scientifique et vulgarisée pour deux raisons. Tout d'abord parce que cela suffit à améliorer considérablement la recherche documentaire : dans les domaines médicaux peuvent être distinguées deux catégories principales d'acteurs, les spécialistes (docteurs, chercheurs) et les autres (patients, journalistes). De plus, des collections importantes de documents sont nécessaires pour appliquer un traitement automatique, ce qui s'obtient plus facilement en se limitant à deux classes. Le type de discours correspond à deux catégories de communication : les documents écrits pour des spécialistes et les documents écrits pour le grand public.

2.2 Sélection et classification des documents

Notre corpus comparable est composé de documents extraits du Web, collectés selon différentes méthodes : moteurs de recherche nationaux français et japonais, portails spécialisés et collections de liens. Les deux premières méthodes nécessitent des mots-clés, nous avons utilisé « diabète », « alimentation », « obésité » pour le français et « 糖尿病 », « 肥満 » (diabète, obésité) pour le japonais. Ces mots-clés ont été étendus avec des synonymes (p. ex. : *nutrition, surpoids, graisse...*) et des termes sémantiquement liés (p. ex. : *insuline, DNID, glucose, calories...*). Les documents ont ensuite été manuellement sélectionnés par des locuteurs natifs de chaque langue et classés selon leur type de discours. La classification manuelle est effectuée selon les heuristiques suivantes :

- un document scientifique est écrit par des spécialistes, à destination de spécialistes ;
- distinguons deux niveaux de vulgarisation : les documents écrits par des spécialistes à destination du grand public et les documents écrits par le grand public. Nous avons privilégié les documents écrits par des spécialistes, généralement plus riches en contenu et en vocabulaire.

Notre classification manuelle est de plus basée sur certains indices empiriques : le site d'origine, le vocabulaire, etc. Nous n'avons pas retenu les documents pour lesquels le type de discours était ambigu, l'auteur étant difficilement identifiable.

2.3 Constitution du corpus et caractéristiques

Les principales caractéristiques du corpus figurent dans le tableau 1.

| | Français | | Japonais | |
|-------------------------------|----------|---------|----------|---------|
| | SC | VU | SC | VU |
| N ^{bre} de documents | 65 | 183 | 119 | 419 |
| N ^{bre} de mots | 425 781 | 267 885 | 234 857 | 572 430 |

Tableau 1 : Caractéristiques du corpus (SC: discours scientifique, VU: discours vulgarisé)

Ce corpus est composé de plus de 1,5 million de mots dans deux langues. Le nombre de mots est évalué à l'aide des espaces et ponctuations entre les mots pour le français et calculé avec les étiquettes ChaSen pour le japonais.

3 Analyse de la modalité

Dans cette section, nous présentons l'analyse de la modalité dans le discours spécialisé médical en français et japonais. Notre but est de confirmer la validité de l'application à notre corpus des deux théories de la modalité présentées ci-dessus. Une première analyse manuelle doit guider l'automatisation de la reconnaissance des critères de modalité.

3.1 Analyse linguistique : d'un processus manuel à un processus automatique

Afin d'observer la modalité et d'adapter nos deux théories, nous avons extrait de notre corpus un ensemble de 5 documents par langue et par type de discours (en moyenne 30 phrases par document) de façon aléatoire afin de conserver une certaine représentativité du corpus. Nous avons exclu les documents trop courts ou dépourvus de phrases.

Nous avons listé les critères de chacune des modalités présentées dans la section 2. La modalité locutive se base principalement sur des verbes, par exemple la modalité de suggestion est exprimée par les verbes « suggérer », « conseiller » en français et « 勧める » (recommander) pour le japonais. Quant à la modalité *Irrealis*, les critères sont des formes infléchies des mots ainsi que les auxiliaires modaux pour le français (ex. : suffixes du conditionnel) et les formes verbales pour le japonais (auxiliaire modal « かも知れない » (pouvoir)). Pour chaque critère de la modalité, un ensemble de mots clés a été créé afin de collecter les occurrences ensuite manuellement filtrées.

Le but de l'analyse automatique est d'étendre cette méthode sur tout le corpus, en ôtant toute intervention humaine. À chaque critère doit donc correspondre

une liste de marqueurs la plus complète possible, afin de collecter un maximum d'occurrences. Une analyse automatique ne pouvant donner des résultats parfaits, nous souhaitons maximiser le nombre d'occurrences collectées tout en minimisant le nombre de faux exemples.

Les deux théories de la modalité sont principalement basées sur des catégories grammaticales et verbes qui peuvent être représentés sous la forme de patrons lexico-syntaxiques. Des étiqueteurs et lemmatiseurs sont appliqués sur notre corpus : les étiqueteurs Brill (Brill 1993) et le lemmatiseur Flemm (Namer 2000) pour le français, l'étiqueteur ChaSen (Matsumoto 1999) pour le japonais.

Deux méthodes nous ont permis d'adapter les listes de marqueurs à l'automatisation :

- l'extension des marqueurs : ajout de synonymes et termes sémantiquement liés ;
- l'ajout de marqueurs utilisant des étiquettes (en particulier pour les verbes conjugués).

L'absence d'occurrences pour un critère lors de l'analyse manuelle ne peut être généralisée au corpus, nous avons donc automatisé la reconnaissance de tous les critères nous paraissant pertinents. Les marqueurs pour chacun des critères de modalité sont basés sur :

- de simples mots-clés (p. ex. : *Monsieur, Madame*, « 皆様 », « 皆さん » pour la modalité de l'interjection) ;
- différentes formes de verbes (p. ex. : verbes conjugués au futur pour la modalité du futur) ;
- des idiomes contenant verbes, adjectifs, adverbes (p. ex. : *être certain de* pour la modalité d'opinion) ;
- des particules, en particulier en japonais (p. ex. : verbe + *こと* pour la modalité d'injonction) ;
- etc. (voir (Nakao et al., 2009) pour plus de détails)

Par exemple, la modalité du jugement est détectée en utilisant le verbe *approuver* conjugué à la première personne, ce qui correspond à l'étiquette « VCJ : 1p : {s|p} » avec Brill (VCJ pour verbe conjugué, 1p première personne, s|p singulier ou pluriel). Avec le lemmatiseur, nous pouvons en français obtenir la forme infinitive des verbes, le verbe *approuver* conjugué à la première personne peut donc être trouvé avec l'expression suivante : « /VCJ : 1p : {s|p}(: .*)/approuver ».

3.2 Résultats

Les tableaux 2 et 3 présentent les résultats de l'analyse automatique sur notre corpus. Pour chacun des critères des deux théories et chaque type de discours, nous avons compté le nombre moyen d'occurrences trouvées dans les documents du corpus. Le tableau 2 présente les résultats de la modalité locutive. Bien que les

résultats soient assez faibles, le nombre moyen d'occurrences est plus important dans les documents scientifiques que dans les documents vulgarisés, quelle que soit la langue. Le nombre moyen d'occurrences est plus important en français qu'en japonais, ceci peut être expliqué par l'importance des pronoms, qui n'existent pas en japonais.

Un critère est dit discriminant s'il est fréquent et qu'il permet de distinguer les deux types de discours.

| | FR | | JP | |
|-------------------------|--------|-------|-------|--------|
| | SC | VU | SC | VU |
| ACTE ALLOCUTIF | | | | |
| Interpellation | 0.17 | 0.045 | 0.98 | 0.48 |
| Injonction | 1.69 | 2.23 | 23.7 | 9.52 |
| Permission | 2.12 | 0.54 | 0.19 | 1.16 |
| Avertissement | 0.75 | 1.76 | 0 | 0 |
| Jugement | 0.12 | 0.096 | 1.31 | 0.93 |
| Suggestion | 5.82 | 0.86 | 0.06 | 0.07 |
| Proposition | 2.46 | 0.49 | 0.43 | 1.01 |
| Interrogation | 21.73 | 16.08 | 1.39 | 1.55 |
| Requête | 0.69 | 0.16 | 0.19 | 0.21 |
| ACTE ÉLOCUTIF | | | | |
| Constat | 35.01 | 15.6 | 0 | 0.005 |
| Savoir/ignorance | 4.85 | 3.34 | 0.21 | 0.33 |
| Opinion | 5.54 | 4.19 | 1.63 | 0.78 |
| Estimation | 47.24 | 18.66 | 1.09 | 0.65 |
| Obligation | 0.24 | 0.27 | 0.06 | 0.03 |
| Possibilité | 13.86 | 7.51 | 16.47 | 6.15 |
| Vouloir | 0.54 | 0.33 | 0.83 | 0.91 |
| Promesse | 0 | 0 | 29.33 | 8.35 |
| Acceptation/refus | 0.49 | 0.17 | 0.1 | 0.2 |
| Accord/désaccord | 10.46 | 5.51 | 0.07 | 0.0096 |
| Déclaration | 2.18 | 0.19 | 0.14 | 0.033 |
| TOTAL | 155.96 | 78.03 | 78.18 | 32.38 |

Tableau 2 : Résultats de la modalité locutive sur le corpus (nombre moyen de marqueurs par document)

Les critères les plus discriminants sont les modalités de l'injonction, de la possibilité et de la promesse en français. De plus, de ces résultats apparaissent quelques informations intéressantes : par exemple, les modalités d'injonction et de promesse sont bien plus fréquentes en japonais qu'en français. Au contraire, les résultats de la modalité de l'estimation et du constat sont supérieurs en français.

Les résultats de la modalité *irrealis* se trouvent dans le tableau 3. Une fois de plus, les occurrences sont faibles pour les deux langues. Dans la dernière ligne du tableau se

trouvent les sommes des nombres moyens d'occurrences de chacune des modalités. La différence des sommes entre chaque type de discours est sensiblement la même pour chacune des deux langues. Bien que les résultats soient comparables pour chaque langue, les critères les plus discriminants sont les verbes de manipulation et les auxiliaires modaux pour le français et les verbes modaux pour le japonais. Certains critères n'obtiennent quasiment aucune occurrence, par exemple les questions oui/non ou les propositions temporelles. Nous observons que certains critères sont trop spécifiques à une langue, par exemple, le conditionnel irréel donne de bien meilleurs résultats en français (et quasiment nuls pour le japonais).

Cette analyse automatique montre que ces modalités sont pertinentes pour distinguer les deux types de discours. Les critères de la modalité locutive sont cependant plus fréquents dans les documents que ceux de la modalité *irrealis*.

| | | FR | | JP | |
|--------------------------|-------------------------------|-------|-------|-------|-------|
| | | SC | VU | SC | VU |
| Temps futur | | 5.22 | 1.5 | 1.07 | 1.09 |
| Adverbes modaux | | 6.54 | 2.51 | 0.21 | 0.05 |
| Compléments de verbes | De modalité | 0.27 | 0.011 | 7.97 | 3.83 |
| | De manipulation | 9.55 | 2.33 | 27.73 | 11.08 |
| | De perception et cognition | 3.88 | 3.76 | 2.81 | 3.02 |
| Discours non déclaratif | Requête | 0.06 | 3 006 | 0.21 | 1.78 |
| | Ordre (1 ^{re} pers.) | 0 | 0 | 0.4 | 2.23 |
| | Ordre (3 ^e pers.) | 0.37 | 0.51 | 0 | 0.002 |
| | Questions oui/non | 0 | 0 | 0 | 0.03 |
| Propositions adverbiales | Propositions temporelles | 0 | 0.04 | 0 | 0 |
| | Conditionnel simple | 0.64 | 0.81 | 0.57 | 1.23 |
| | Conditionnel subjonctif | 0.47 | 0.63 | 0.14 | 0.011 |
| | Conditionnel irréel | 0 | 0.05 | 6.04 | 4.05 |
| Auxiliaires modaux | | 26.31 | 12.57 | 6.21 | 3.41 |
| TOTAL | | 53.31 | 27.73 | 53.36 | 31.81 |

Tableau 3 : Résultats des critères de la modalité *irrealis* sur le corpus (nombre moyen de marqueurs par document)

3.3 Discussion

À première vue, nos résultats, bien que faibles, paraissent efficaces pour distinguer les deux types de discours, scientifique et vulgarisé, dans des documents médicaux. En ce sens, nous pouvons affirmer que les théories de la modalité choisies sont adaptées aux langues de spécialité en français et en japonais. Nous pouvons cependant discuter ces résultats selon deux phénomènes : le bruit (détection de faux exemples) et le silence (non-détection d'exemples).

Plusieurs raisons peuvent expliquer le bruit :

- l'ambiguïté des marqueurs : par exemple, le verbe *pouvoir* en français est supposé indiquer la modalité d'autorisation, il peut pourtant avoir d'autres sens

[16] Pouvez-vous nous parler des problèmes de vision associés au diabète ?

- l'ambiguïté des étiquettes morphosyntaxiques : quand certains mots correspondent à plusieurs formes, elles sont toutes indiquées dans l'étiquette. Par exemple, en français, l'indicatif du présent et l'impératif ont certaines conjugaisons communes, ce qui est indiqué dans l'étiquette :

[17] Il est conseillé de bien choisir les graisses que vous consommez.

[18] Consommez-le en quantité contrôlée et régulière ;

- l'influence de la position : en japonais, la position des marqueurs indique sa nature modale. Par exemple, dans la phrase :

[19] « 2.2 節で述べたように、発現ピーク時間での血糖値を予測できれば、適切なインスリン投与量、すなわち設定量で良いのか、増やすべきか減らすべきかを適切に判断できる » (comme nous le mentionnons dans la section 2.2, si le pic de taux de sucre dans le sang peut être estimé, nous pouvons déterminer la quantité appropriée d'insuline à injecter, c'est-à-dire juger si la quantité fixée doit être augmentée/diminuée)

le marqueur « verbe + ように » n'indique pas la modalité. Toutefois, ce même marqueur trouvé à la fin d'une phrase avec un verbe au présent indique la modalité d'injonction :

[20] « 決められた用量を守るように。 », (vous devez respecter la dose prescrite).

Le bruit est souvent introduit par la modalité locutive car ses marqueurs sont simples (verbes ou mots-clés) ou basés sur des étiquettes morphosyntaxiques potentiellement ambiguës. Le silence est introduit lorsque les marqueurs d'un critère ne couvrent pas toutes les occurrences. Quand les marqueurs sont des mots-clés, nous avons cherché à faire une liste exhaustive de marqueurs, plus susceptible d'introduire du bruit que du silence. Cependant, lorsque les marqueurs sont plus précis et représentent des phénomènes plus complexes, ils introduisent du silence. Par exemple, le conditionnel est un critère plus complexe : en français, il est

nécessaire de détecter un *si*, suivi d'un verbe conjugué dans une forme particulière, puis suivi par un autre verbe sous une forme différente (p. ex. : *si* + présent + présent).

Une autre cause du silence est la présence de marqueurs implicites. Dans ce cas, nous pouvons croire que des marqueurs très spécifiques peuvent être efficaces, mais cette solution ne l'est pas toujours. Par exemple, en japonais, « 明らかにしてきた » exprime la modalité du constat, mais ce marqueur est trop spécifique et introduit du bruit. Le silence est plus présent avec la modalité *irrealis*, car les critères sont basés sur des unités plus complexes.

Conclusion

Dans cet article, nous avons évalué si l'analyse de la modalité dans des documents médicaux pouvait être pertinente afin de distinguer les deux types de discours scientifique et vulgarisé. En nous basant sur deux théories de la modalité définies sur la langue générale, nous les avons adaptées au domaine médical en français et japonais et à une analyse automatique. Nous avons de plus évalué les résultats de cette analyse automatique par rapport aux deux types de discours. Cette étude a été menée en deux étapes. Nous avons dans un premier temps examiné si ces deux théories définies sur la langue générale pouvaient être adaptées aux langues de spécialité. La modalité locutive semble, en théorie, adaptée aux langues de spécialité. Dans un discours supposé neutre, la présence, même rare, du locuteur ou de l'interlocuteur donne des informations sur le type de discours. Cette supposition s'est confirmée avec l'analyse automatique, cette modalité est pertinente pour distinguer ces types de discours dans les langues de spécialité. La théorie *irrealis* révèle la subjectivité qui, en théorie, nous semble pertinente. Son expression indique la position du locuteur quant à son discours. Même si l'analyse automatique a mis en évidence certains problèmes avec cette théorie, les résultats montrent qu'elle est adaptée aux langues de spécialité. Elle paraît cependant moins discriminante que la modalité locutive.

Nous avons dans un second temps évalué si ces théories étaient adaptées aux langues française et japonaise. En français, la modalité locutive donne de bons résultats. Comme cette théorie a été définie pour le français, cela semble normal. La théorie *irrealis* donne de moins bons résultats, ce qui peut être dû à l'automatisation. Manuellement, ces deux théories semblent pertinentes, mais la théorie *irrealis* passe plus difficilement à l'échelle du corpus.

En résumé, la modalité est pertinente pour distinguer les types de discours dans un domaine spécialisé. Les deux théories sont applicables aux langues française et japonaise, bien qu'elles aient été définies sur une seule langue. Nous avons montré que la modalité *irrealis* n'était pas forcément bien adaptée à une automatisation, contrairement à la modalité locutive. L'apport d'une analyse de la modalité sous la forme de classification automatique de documents spécialisés semble très

intéressante pour des système de recherche d'information dans les domaines spécialisés.

Bibliographie

- Bowker (L.) et Pearson (J.), 2002 : *Working with Specialized Language: A Practical Guide to Using Corpora*, London/New York: Routledge.
- Brill (E.), 1993 : *A Corpus-based Approach to Language Learning*, PhD thesis, University of Pennsylvania, USA.
- Charaudeau (P.), 1992 : *Grammaire du sens et de l'expression*, Paris: Hachette.
- Chung (S.) et Timberlake (A.), 1985 : « Tense, aspect, and mood », dans *Language typology and syntactic description*, III, Cambridge University Press.
- Elhadad (N.) et Sutaria (K.), 2007 : « Mining a lexicon technical terms and lay equivalents », dans *Proceedings of ACL BioNLP Workshop*, Prague, Czech Republic, p. 49-56.
- Givón (T.), 1994 : « Irrealis subjunctive », *Studies in language* 18 (2), p. 265-337.
- Habert (B.), Grabar (N.), Jacquemart (P.) et Zweigenbaum (P.), 2001 : « Building a text corpus for representing the variety of medical language », dans Rayson (P.), Wilson (A.), McEnery (T.), Hardie (A.) et Khoja (S.), éd., *Proceedings of Corpus Linguistics 2001*, 245-254, Lancaster, UK.
- Le Querler (N.), 1996 : *Typologie des modalités*, Caen: Presses universitaires de Caen.
- Malrieu (D.) et Rastier (F.), 2002 : « Genres et variations morphosyntaxiques », dans *Traitement Automatique des Langues (TAL)* 42 (2), p. 548-577.
- Masuoka (T.) et Takubo (Y.), 1992 : *Kiso Nihongo Bunpou*, Tokyo: Kuroshio-syuppan (japonais).
- Matsumoto (Y.), Kitauchi (A.), Yamashita (T.) et Hirano (Y.), 1999 : *Japanese Morphological Analysis System ChaSen 2.0 Users Manual*, technical report, Nara Institute of Science and Technology (NAIST).
- Miyazaki (K.), Adachi (T.), Noda (H.) et Sinano (T.), 2002 : *Modalitii*, Tokyo: Kuroshio-syuppan (japonais).
- Morin (E.), Daille (B.), Takeuchi (K.) et Kageura (K.), 2007 : « Bilingual Terminology Mining — Using Brain, not brawn comparable corpora », dans *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL'07)*, Prague, Czech Republic, p. 664-671.
- Nakao (Y.), Goeuriot (L.) et Daille (B.), 2010 : « Multilingual Modalities for Specialized Languages », dans *Terminology*, 16 (1), p. 51-76.
- Namer (F.), 2000 : « Flemm: Un analyseur Flexionnel de Français à base de règles », dans Jacquemin (C.), éd., *Traitement automatique des langues pour la recherche d'information*, Christian Jacquemin (éd.). Paris: Hermès, p. 523-547.
- Nomura (T.), 2003 : « Modalitii bunrui no keisiki », dans *Kokugogaku* 54 (1), p. 17-31.
- Palmer (F. R.), 1986 : *Mood and Modality*, Cambridge: Cambridge University Press.
- Pearson (J.), 1998 : *Terms in Context*, Amsterdam/Philadelphia: John Benjamins.
- Pecman (M.), 2007 : « Approche onomasiologique de la langue scientifique générale », dans *Revue française de linguistique appliquée*, XII (2), p. 79-96.

- Sionis (C.), 2002 : « Quelques spécificités de la modalisation dans le discours scientifique », dans *ASp* 35-36, p. 1-13.
- Tamachi (M.), 2005 : « Nihongo to Cyuugokugo no modalitii no taisyou kenkyuu », dans *Takamatsu Daigaku Kiyou*, n° 14, p. 17-54.
- Vinot (R.), Grabar (N.), Valette (M.), 2003 : « Application d'algorithmes de classification automatique pour la détection des contenus racistes sur l'Internet », dans *Actes du colloque TALN 2003*, 11-14 June 2003, Batz sur Mer, p. 257-284.

Emmanuel Cartier,
Laboratoire LDI UMR 7187,
Université Paris 13, CNRS,
Villetaneuse,
France,
ecartier@LDI.univ-paris13.fr.

Repérage automatique des contextes définitoires dans les textes

Résumé

Cet article tente de poser des jalons théoriques sur une information centrale en lexicographie et en sémantique, l'information définitoire. Nous tentons tout d'abord d'en délimiter le champ conceptuel, en nous aidant des apports à la fois de la philosophie mais également de la linguistique théorique ; nous aboutissons ainsi à un modèle opérationnel de l'information définitoire, telle qu'elle se rencontre dans les textes réels. La seconde partie expose les grands marqueurs linguistiques utilisables en français pour exprimer ce type d'informations ; nous détaillons dans ce cadre à la fois les schémas syntactico-sémantiques propres aux expressions définitoires, mais également les contraintes de mises en discours qui affectent ces schémas canoniques. Enfin, nous évoquons une première expérience d'automatisation des informations définitoires dans les textes, à partir des descriptions préalablement explicitées. Ce travail est une continuation du travail de thèse de l'auteur, actuellement développé dans le laboratoire LDI.

Mots-clés : définition ; contexte définitoire ; relation sémantique ; traitement automatique des langues ; lexicographie ; usage.

Introduction

La notion de définition est une notion centrale en lexicologie, car dans cet objet textuel se ramassent les informations les plus essentielles du sens des mots et des catégories d'objets qu'ils désignent. Sur cet objet repose donc une très grande part de la validité et de l'utilité des dictionnaires papier et électroniques. Dans cet article, nous voudrions poser quelques jalons permettant de repérer automatiquement dans les textes ce que nous appellerons des *contextes définitoires*, à savoir des phrases dans lesquels on retrouve ce type d'information, sans pour autant être situés nécessairement dans un contexte dictionnaire ; notre intérêt ici est donc d'étudier les formes linguistiques de la définition dans les textes, reflet de la pratique définitoire naturelle, scientifique ou technique actuelle.

Évidemment, les contextes définitoires seront plus nombreux dans les documents de type scientifique et pédagogique, ce qui justifiera le choix d'un corpus de travail issu de l'encyclopédie *Universalis*.

L'objectif de ce travail est de décrire les formes linguistiques de la définition dans les textes, des formes les plus explicitement définitoires aux formes les plus implicites et les plus discrètes. L'objectif second est l'identification automatique de ces contextes définitoires, par le biais d'une description formalisée des schémas syntactico-sémantiques propres aux expressions définitoires.

Notre présentation comprendra trois grandes sections : une première section délimitera l'information définitoire en explicitant une conception basée sur le travail de thèse de l'auteur (Cartier 2004) ; une seconde partie, plus proprement linguistique, s'attachera à décrire les expressions définitoires du français ; sans prétendre dans le cadre cet article couvrir l'ensemble des expressions, nous chercherons à illustrer les grandes difficultés incontournables dans tout travail de description formalisée, en vue d'une implémentation informatique. Enfin, nous présenterons quelques résultats initiaux de l'implémentation des descriptions sur un corpus issu de l'encyclopédie *Universalis*.

1 Définition et informations définitoires

Nous partirons d'une conception opérationnelle des expressions définitoires résumée dans le schéma suivant (pour le cheminement pour y parvenir, voir (Cartier 2004)):

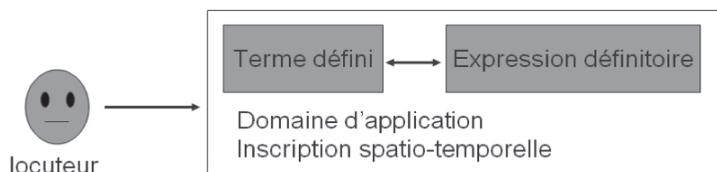


Figure 1

Dans ce schéma, une expression définitoire hérite tout d'abord des propriétés de toute proposition énoncée, à savoir d'être prise en charge énonciativement, ce qui est explicité ici par un locuteur et une flèche (représentant le type de prise en charge énonciative) pointée vers une « proposition ». De cette propriété découle que toute définition est par sa nature même une définition de type prescriptif, en tout cas, en dépit des visées affirmées des locuteurs, un locuteur ne peut s'abstraire de la situation linguistique d'énonciation, qui rend tout énoncé définitoire subjectif et résultat d'une interprétation individuelle. Il en est de même par exemple des définitions lexicographiques qu'on rencontre dans les dictionnaires, même si, une fois la subjectivité intrinsèque et indépassable reconnue et acceptée, il faut rendre compte des visées définitoires du locuteur : définition de choses, de mots. Il faut également rendre compte des définitions à visée prescriptive (par exemple :

j'appellerai *BeauRuche* un rassemblement d'abeilles en cours de nidification), les définitions de l'usage ou de mots (Temps [...] 1. Milieu indéfini où paraissent se dérouler irréversiblement les existences dans leur changement, les événements et les phénomènes dans leur succession [...] (*Petit Robert*, édition 1986)), les définitions de choses (Le temps est le nombre du mouvement, selon l'avant et l'après. (Aristote, *Physique*, IV, 10-14)). Au final, l'usage définitoire « naturel » est finalement celui qui parle des choses.

S'agissant du « contenu » de la proposition définitoire, nous proposons de la décomposer en trois composantes : dans toute sa généralité, une définition n'est rien d'autre qu'une *identification* d'un terme défini à une expression définissante.

1. Nous parlons d'« identification » pour éviter le piège de la définition d'un point de vue philosophique ; le débat philosophique, qui ne nous concerne pas ici, s'intéresse au contenu de la définition, à ce qui *doit* être dit dans une définition. Notre intérêt est de rendre compte des définitions effectives, qu'elles soient proférées par l'individu compétent dans une langue ou par le locuteur spécialiste d'une discipline ; dans les deux cas, il reste une compétence définitoire qui permet à tout un chacun de tenter de dire ce que l'on veut dire en employant tel ou tel mot, telle ou telle expression. L'usage des mots, le sens des mots, est nécessairement accessible à tous. Nous préférons donc parler d'identification, pour expliciter d'une part cette compétence sans ses connotations philosophiques, mais également pour expliciter l'autre propriété essentielle de la définition : le fait qu'elle doit dire l'essentiel du sens du terme défini, et non pas une propriété accidentelle ; du point de vue lexicographique, l'expression définitoire est donc le creuset principal de la sémantique stable et profonde des signes linguistiques. Le repérage de telles expressions est d'autant plus intéressant lorsqu'elles sont proférées par des personnes dont l'objectif n'est pas de faire des définitions lexicographiques car, en quelque sorte, elles vont dire plus naïvement que le lexicographe, ce qu'elles comprennent par tel signe, et comment elles le comprennent.
2. un premier élément de cette relation d'identification est le *terme défini*, élément qui permet d'identifier les objets ou la catégorie d'objets désignés par le signe linguistique ;
3. le second élément est de type discursif, et nous l'appellerons « expression définissante » ; nous en verrons dans la suite un certain nombre de formes.

Il reste encore deux éléments intrinsèques à toute expression définitoire : tout d'abord son inscription spatio-temporelle qui prend racine dans la prise en charge énonciative de la proposition, qui peut donner lieu à différents effets de polyphonie (*Aristote pensait que le temps devait se concevoir comme.../Selon Newton, le temps est le principe de.../Aristote s'est cantonné à définir le temps comme...*). Ces effets de polyphonie permettent de restreindre la portée des définitions à certaines périodes ou à certains espaces géographiques (*Pour les Jurassiens, le terme chat désigne une partie de cartes jouée à l'occasion d'un décès*).

Enfin, toute définition comme toute proposition énoncée dispose d'un domaine d'application qui peut être implicite, auquel cas il s'agira du domaine général, ou bien d'un domaine spécifique qui devra alors être explicité : *pour les chasseurs, l'échasse désigne une manière spécifique de chasser la perdrix/ Les méthodistes pensent que le temps est.../En mathématiques, la transitivité algébrique se définit comme...*

2 Expressions définitives du français contemporain

Les différentes composantes d'une expression définitive doivent faire l'objet d'une étude séparée. Dans le cadre de cet article, nous ne donnerons que quelques indications générales, renvoyant à (Cartier 2004) pour le détail. Dès l'abord, nous distinguerons entre les expressions définitives complètes et les expressions définitives partielles. Les premières sont les définitions proprement dites, mais si nous voulons rendre compte de l'ensemble du champ d'expression définitive, nous devons également rendre compte des expressions rendant compte de propriétés/relations définitives : la catégorisation, la spécification, la synonymie, la relation méronymique/holonymique, la causalité, la téléonomie et l'attribution essentielle font ainsi partie des relations à caractère définitive. Toute tentative de description exhaustive devra donc inclure ces expressions dites partielles.

2.1 Expressions linguistiques de l'identification complète

| Valeur sémantique | Marqueurs et formulation(s) |
|---|---|
| Identification pure | <i>Être</i> (voir étude particulière) SN ₁ , c'est-à-dire SN ₂ SN ₁ , ou (Adv) SN ₂ SN ₁ (SN ₂) SN ₁ , SN ₂ , SN ₁ – SN ₂ - SN ₁ = SN ₂ SN ₁ : SN ₂ |
| Identification pure + expression de la subjectivité | Les marqueurs précédents + adverbiaux, incises, etc. <i>s'interpréter comme, considérer comme</i> <i>en tant que/comme</i> Verbes de pensée/de parole (<i>penser, croire, douter, etc.</i>) <i>que</i> + marqueurs identification pure |
| Identification pure + inscription temporelle/aspectuelle | <i>rester, devenir, évoluer vers, changer en</i> marques aspecto-temporelles (<i>à l'époque de.../autrefois,</i> marques verbales) + identification pure |
| Identification pure + focalisation sur le type d'identification | <i>comme, être similaire à, ressembler à, assimiler à, paraître</i> <i>comme, être identique à...</i> |

Tableau 1

Le tableau 1 synthétise les principales formes de l'expression définitive complète : l'identification pure emploie principalement en français la médiation de *être* (avec de nombreuses contraintes sur la forme de ces arguments), deux marqueurs

conjonctionnels (*c'est-à-dire* ou *ou*), des signes de ponctuation (apposition, parenthèse, juxtaposition par : ou autre signe réglé par un usage spécifique de type lexicographique). On retiendra de cette série que le français utilise une vaste gamme de signes linguistiques, qui partent de *être* (dans son sens le plus pauvre, si l'on peut dire), à la simple juxtaposition des deux composantes d'une définition, comme le font de très nombreuses langues, simplement marquées par des signes de ponctuation.

À partir de là, les expressions se répartissent selon un apport sémantique qui ressortit à trois composantes du sens : la subjectivité, la temporalité-aspectualité et la focalisation sur le type d'identification. Chaque fois, il est à noter que les expressions sont soit spécifiques soit utilisant les marqueurs principaux dotés de marqueurs additionnels (par exemple X interpréter Y comme Z versus X penser que Y être un Z).

2.2 Expressions spécifiques de la définition : définir

Le champ sémantique restreint de la définition s'exprime au moyen du radical *défin-* qui se matérialise en français dans *définition*, *défini*, et (*se*) *définir*. Pour nous concentrer sur les expressions verbales, voici l'organisation syntaxico-sémantique du verbe (tableau 2) :

| Valeur sémantique | Schémas syntaxico-sémantiques | Exemples |
|-------------------|---|--|
| Définition | SNLoc <i>définir</i> SN ₁ (<i>comme</i> :)? SN ₂ | Le même auteur définit, par opposition, le conflit comme « une situation dans laquelle les parties sont conscientes de l'incompatibilité de positions futures potentielles et dans laquelle chaque partie désire occuper une position qui est incompatible avec les désirs de l'autre ». |
| | SN ₁ <i>être défini</i> (<i>comme</i> :)? SN ₂ (<i>(selon pour d'après par)</i> SNLoc)? | La molécule est alors définie comme « la Plus Petite quantité dans laquelle, par combinaison ou par substitution, on puisse faire entrer un nombre irréductible d'atomes d'un élément quelconque ». |
| | SN ₁ <i>se définir</i> (<i>comme</i> :)? SN ₂ (<i>(selon pour d'après par)</i> SNLoc)? | La molécule se définit comme « la plus petite quantité dans laquelle, par combinaison ou par substitution, on puisse faire entrer un nombre irréductible d'atomes d'un élément quelconque ». |
| Caractérisation | SNLoc <i>définir</i> SN ₁ <i>par</i> SN ₂ | Elle définit la concentration par son effet, la création d'un pouvoir de contrôle d'une entreprise sur une autre. |
| | SN ₁ <i>se définir</i> (<i>par</i> SN ₂ <i>comme</i> SAdj ₂) SN ₁ <i>être défini</i> (<i>par</i> SN ₂ <i>comme</i> SAdj ₂) SN ₂ <i>définir</i> SN ₁ | Le chômage « au sens du B.I.T. » n'est défini ni par la déclaration spontanée, ni par la situation juridique des individus, mais par un jeu de critères mis en rouvre dans un questionnaire détaillé, élaboré par les instituts statistiques nationaux. |

| | | |
|-------------------------------|---|---|
| Détermination/ Implication | SN ₁ <i>définir</i> SN ₂ | Un autre indicateur définit les différences entre ces régions: le revenu par tête varie énormément entre le Nord et le Sud. |
| | SN _{Loc} <i>Définir</i> SN ₂ | Le concept de tribu est donc « en crise », et il y a double urgence, théorique et pratique, à remonter aux racines du mal qui l'atteint et à le redéfinir pour en faire la critique et en estimer la portée réelle. |
| | SN ₁ <i>se définir</i> (+ Circonstant) SN ₁ <i>être défini</i> (+ Circonstant) | Les valeurs d'échange globales peuvent être définies négativement ou positivement selon les objectifs. |

Tableau 2

La structure verbale du verbe *définir* permet l'explicitation de tous les participants à l'acte de définition: une personne prenant en charge la définition (SN_{Loc}), un terme défini (SN₁) et un syntagme définissant (SN₂). Dans les schémas 1 et 2, l'explicitation du locuteur est facultative, sous forme d'un circonstanciel. Par contre, dans le schéma 3, les trois actants sont présents.

La valeur de caractérisation se distingue principalement de la valeur définitoire par la structure du complément SN₂ ou SADJ₂: il est introduit par la préposition *par* qui indique la propriété définitoire mise en valeur, ou introduit par *comme* suivi d'une structure adjectivale.

La valeur de détermination/implication se distingue des valeurs définitoires principalement par deux propriétés: d'une part, le SN₁ n'est plus ressenti comme étant un concept propre, un terme défini, et alors la valeur glisse vers une valeur de détermination/implication; d'autre part, lorsque SN₂ est omis, on glisse vers un sens de pré-définition/détermination, qui rappelle le sens étymologique du terme de « délimitation de territoire ».

2.3 Expressions spécifiques de la définition : *s'appeler, signifier, désigner, dénoter*

Un champ connexe concerne l'expression de la dénomination/signification/dénotation concrétisée par les marqueurs verbaux principaux suivants: *s'appeler, signifier, désigner, dénoter*. Ces verbes, selon (Rey-Debove 1978) permettent de rendre compte de trois types d'identification à partir de l'identification « vide », au moyen de la copule *être*: l'identification insistant sur le nom (*s'appeler*), insistant sur l'objet mondain (*désigner*), insistant sur le sens (*signifier*). On peut les appeler *marqueurs spécialisés*, car ils expriment principalement une définition, ou l'un de ses éléments.

2.4 Des expressions définitoires partielles

Les expressions définitoires partielles représentent le second groupe d'expressions à décrire. Nous avons opté pour une classification reposant sur les types

d'informations définitoires explicitées, ce qui nous donne autant de groupes que d'informations définitoires. Voici, à titre illustratif, les expressions linguistiques du français utilisables pour les expressions de catégorisations.

a) Coréférence anaphorique et nominale: le premier moyen d'expressions de la catégorisation en français ressortit au phénomène de coréférence anaphorique et nominale; ce moyen est sans doute le plus intéressant, car il s'agit d'une expression naturelle de la catégorisation. On le rencontre dans toute communication linguistique, sous des formes extrêmement diverses:

- recatégorisation par descriptions définies: «... Dubaï... En 2008, l'émirat avait attiré à lui seul... » (*Le Figaro*, 27/11/2009)
- catégorisation intra-nominale dans des syntagmes nominaux composés d'un nom commun suivi d'un nom propre: « le directeur général du club Claude Machin... »
- appositions: « les chemins de fer français, infrastructure mise en place dès... »
- recatégorisation plus complexe: par exemple sur le quai d'une gare TGV: « après un court arrêt, le train sera déplacé de quelques mètres. Les voyageurs sont priés d'attendre la fin de la manœuvre pour monter dans le train. »

b) Partie initiale de l'information définitoire dans la définition complète: bien évidemment, l'information catégorielle est une des composantes d'une définition complète: le Nox est un gaz à effet de serre qui...

c) Expressions spécialisées: enfin, le français dispose de toute une série d'expressions spécialisées dans ce type d'expression, et notamment:

- ... Relateur *définitoire* + Déterminant défini + catégoriseur *générique* (*espèce, classe, genre, etc.*) + *catégorie*: « Le Nox est une espèce de gaz »
- Préposition catégorisante (*parmi, à l'intérieur des, dans...*) + catégorie, espèce: « Parmi les gaz, le Nox... »
- Expressions verbales spécifiques: X (*faire partie de, appartenir...*) à (le genre, la catégorie) des Y.

2.5 Du repérage des composantes de la définition : terme défini et expression définissante

L'étude précédente nous permet de repérer globalement les expressions définitoires dans les textes. Il reste ensuite à repérer précisément le terme défini et l'expression définissante. Comment identifier ces deux informations?

2.5.1 Du terme défini

L'élément le plus simple à repérer est apparemment le terme défini. En effet, en français, un terme défini est soit une expression nominale répondant à l'un des

schémas des noms français, soit une expression en mention, marquée typographiquement et/ou par des introducteurs spécifiques (*le mot « effet de serre », « Prendre ses jambes à son cou » signifie...*). La principale difficulté réside donc dans le repérage des 762 structures morpho-syntaxiques des noms composés en français (voir Mathieu-Colas 2009).

2.5.2 De la décomposition de l'expression définissante

L'expression définissante propose des formes morphosyntaxiques en plus grand nombre. Dans la définition complète, l'expression définitoire peut prendre deux formes principales : un groupe nominal complexe ou une proposition. En voici deux exemples :

- [1] Le chômage se définit comme un état de personnes recherchant un emploi et n'en trouvant pas.
- [2] Un nombre est pair si et seulement si il peut être divisé par lui-même et par un autre nombre et que le résultat est un nombre entier.

Ces deux types d'expression sont, par le biais d'une transformation syntaxique, équivalentes :

- [1'] Il y a chômage si et seulement si il y a des personnes recherchant un emploi et n'en trouvant pas.
- [2'] Un nombre pair est un nombre qui peut être divisé par lui-même, le résultat de la division étant un nombre entier.

On constate que dans le cas (2), le concept est défini par des conditions (ici nécessaires et suffisantes) qui fixent la possibilité d'existence du concept. Dans le cas (1), on pose les propriétés nécessaires à l'existence de ce concept. Dans le cas (1), l'expression définitoire comprend toujours une indication de genre suivie d'une ou plusieurs propriété(s) nécessaires ; dans le cas (2), la proposition définitoire indique seulement les propriétés nécessaires, et le membre défini contient une indication de genre suivi du relateur *être* suivi d'une propriété (ici : un *nombre est pair*). On constate donc que, dans le cas (2), il est aisé d'identifier les composantes définitoires, puisqu'elles sont dissociées dans l'expression elle-même (*Genre être Propriété ssi Condition₁, Condition₂, etc.*). Dans le cas (1), par contre, étant donné qu'elles sont combinées, il faut chercher la frontière qui permet de les distinguer, sachant que *le genre précède toujours la ou les propriétés, sous forme nominale*, du moins dans les expressions françaises ; la ou les propriétés sont exprimées par une proposition relative ou circonstancielle. Il reste quelques cas limites, comme *Le Nox est un gaz à effet de serre*, dans laquelle il est quasi impossible de distinguer le genre et les propriétés. Nous aboutissons donc, sans prétendre à l'exhaustivité mais plutôt à la prototypie, à la classification suivante des informations définitoires dans la définition complète (tableau 3) :

| Type spécifique | Exemples |
|-----------------|----------|
|-----------------|----------|

| | |
|---|--|
| (<i>Si, quand, lorsque</i>) + Proposition | Un homme est un menteur s'il dit des choses qui ne sont pas vraies |
| Syntaxme nominal SN + proposition relative | Un menteur est un homme qui dit des choses qui ne sont pas vraies |
| SN + proposition circonstancielle | Un menteur est cet homme, lorsqu'il dit des choses fausses |
| SN + proposition gérondif | Un menteur est un homme disant des choses fausses |
| SN + proposition infinitive | L'athéisme est l'idée de nier l'existence de Dieu |
| SN + PropComplétive | L'athéisme est l'idée que Dieu n'existe pas |

Tableau 3

2.6 Des autres informations dans une expression définitoire

S'agissant des autres informations composant une expression définitoire, il reste : l'expression de la source et de la prise en charge énonciative ; du domaine de définition et de l'inscription aspecto-temporelle. Nous ne détaillerons pas ici ces expressions, rappelant simplement que pour ces informations est en œuvre le principe d'économie voulant par exemple que l'absence d'informations énonciatives explicites donne pour valeur énonciative la valeur actuellement en cours dans le texte, ou bien la valeur source énonciative « je » et la prise en charge neutre « penser, dire » (voir Cartier, 2010 pour le détail de cette analyse).

2.7 Des phénomènes d'intégration textuelle

Une fois les informations définitoires précédentes repérées, nous obtenons une représentation formalisée et exhaustive des expressions canoniques du français. *Canoniques*, car les expressions décrites ne tiennent pas compte des paramètres de mise en discours qui peuvent changer les formes discursives des propositions. Prenons une phrase exemple. Soit le schéma définitoire Terme défini être Genre Propriété distinctive. (Ou : définition (terme Défini, Genre, Propriété(s) distinctives)) ; en partant des termes définis « Nox », du genre « gaz à effet de serre », de la propriété distinctive « nocif », et du relateur *être* (tous ces éléments étant des éléments lexicaux), quelles sont les formes possibles en français de cette expression ?

- Tout d'abord, il y a l'application de règles morphosyntaxiques qui contraignent à la fois la forme des composants lexicaux (par exemple : *nocif* = > *nocive*), mais également l'ordre des composants dans la phrase. Ainsi dans la phrase *Le Nox est un gaz à effet de serre nocif* relèvent de la morphologie les morphèmes pluriel ou singulier des noms et adjectifs et les morphèmes verbaux ; relèvent de la syntaxe l'ordre entre déterminant et contenu sémantique nominal, ainsi que l'ordre SVO (sujet verbe objet). Ces éléments sont propres à chaque langue.
- Il existe aussi un certain nombre de transformations possibles de cette forme canonique : passivation, pronominalisation, interrogation, interro-négation, etc.

- Des phénomènes d'insertion sont possibles entre les constituants essentiels de la proposition : au niveau de la proposition, insertion d'adverbiaux, insertion de particules de négation ; au niveau des constituants, insertion de verbes modaux, etc.
- Des déplacements peuvent avoir lieu, permettant des antépositions, postpositions, ellipses : *Défini comme un gaz à effet de serre nocif, le Nox...*
- Des remplacements (anaphore, cataphore) peuvent également se produire sur chacun des constituants lexicaux.

3 Implémentation informatique des descriptions des expressions définitives

Le travail de description des expressions définitives du français, de par ces caractéristiques de formalité et exhaustivité, peut être implémenté dans un moteur de repérage d'informations textuelles. Nous avons utilisé un outil par ailleurs développé par l'auteur, Textbox. Nous avons, dans une phase préliminaire, pris pour objectif de repérer les contextes définitives canoniques : cette première phase, sur la base des marqueurs définitives et des schémas syntactico-sémantiques attachés à chaque sens, permet de repérer le contexte définitive en désambiguïsant les différentes mentions du marqueur. Les ressources linguistiques utilisées pour mettre en œuvre ce repérage sont de deux types : des classes de mots et des schémas syntactico-sémantiques.

3.1 Classes de mots

Les classes de mots sont des listes de mots à sémantisme équivalent ou synonymique et à fonctionnement syntaxique identique ; nous avons alors défini un total de 56 classes de mots, réparties comme suit : 19 classes de mots outils, 37 classes spécifiques du repérage des expressions définitives. Au total, nous disposons ainsi de 2076 marqueurs utiles. Chacune des classes est décrite, d'une part en termes de ses membres, mais aussi de différents attributs utiles et utilisables ensuite dans les schémas syntactico-sémantiques. Ces attributs descriptifs sont de différentes natures : morphologique (modèle morphologique avec attributs nombre, personne, genre pour les noms et adjectifs, et nombre, personne, temps pour les verbes) ; morphosyntaxique « partie du discours » (nom, adjectif, pronom, déterminant, verbe, etc., avec éventuellement une sous-catégorisation, notamment pour les déterminants) ; sémantique (indication de la langue et de la classe sémantique, par exemple « catégoriseur » pour les noms catégoriseurs du type « espèce, genre, phénomène, objectif », etc., avec éventuellement une sous-catégorisation sémantique, par exemple, pour les catégoriseurs, « espèce, genre, méronymie », etc.).

3.2 Schémas syntactico-sémantiques

Les schémas permettent de décrire les séquences de mots impliqués dans un type d'expression particulier. Par exemple, voici la règle pour repérer quelques expressions définitoires avec *être* (tableau 4) :

| Schéma syntactico-sémantique | Valeur sémantique | Exemple |
|---|---|---|
| SN ₁ (DetDef + SN) être SN (DetIndef + SN ₂ + X ₃ (PropRel PropCirc Sadj)) | define (word = SN ₁ , Gender = SN ₂ , Prop = X ₃) | <i>Les Nox sont des gaz à effet de serre nocifs</i> = > define (word: Nox, gender: gaz à effet de serre, Prop: nocif) |

Tableau 4

3.3 Évaluation initiale du système

Nous avons procédé à une première évaluation du système en lançant le système sur mille phrases issues de l'encyclopédie *Universalis*, avec 79 schémas. Le tableau 5 récapitule les résultats pour chaque jeu de règles.

| Marqueur | Nombre de règles | Précision | Rappel |
|----------------------|------------------|-----------|--------|
| <i>Être</i> | 13 | 87 % | 76 % |
| <i>Définir</i> | 7 | 94 % | 57 % |
| <i>Désigner</i> | 6 | 97 % | 85 % |
| <i>Dénoter</i> | 8 | 98 % | 92 % |
| <i>S'appeler</i> | 5 | 98 % | 95 % |
| <i>Signifier</i> | 7 | 97 % | 97 % |
| <i>Comme</i> | 3 | 85 % | 45 % |
| <i>Ou</i> | 5 | 81 % | 52 % |
| <i>C'est-à-dire</i> | 7 | 87 % | 61 % |
| <i>Autrement dit</i> | 7 | 88 % | 79 % |
| <i>À savoir</i> | 6 | 94 % | 90 % |
| Parenthésage | 3 | 73 % | 68 % |
| Apposition | 2 | 52 % | 61 % |
| Total | 79 | | |

Tableau 5

On peut faire les remarques principales suivantes : 1) les taux de rappel sont plus faibles que les taux de précision : il reste encore du bruit dans le repérage des contextes définitoires ; 2) les marqueurs les plus polysémiques (*être*, *ou*, *comme*, *apposition*,) ont des taux encore insuffisants, dus à leur grande polysémie ; 3) au contraire, les marqueurs les plus explicites (*à savoir*, *s'appeler*, *désigner*, *signifier*) ont des taux très corrects. Pour conclure, nous dirons que ces résultats sont encore

partiels, étant donné que la tâche finale consisterait à repérer les composantes sémantiques d'une relation définitoire, et non pas le seul contexte. Cette implémentation sera faite dans un travail ultérieur.

Conclusion

Le travail présenté dans cet article a un objectif principalement méthodologique : à partir d'une information particulièrement importante dans l'étude de la sémantique des langues, il a visé à expliciter une méthodologie aboutissant à des descriptions linguistiques caractérisées par la précision et l'exhaustivité ; ces deux traits sont assurés par l'implémentabilité des descriptions, et donc la possibilité de vérifier sur corpus la validité des descriptions linguistiques. De ce point de vue, nous pensons avoir réussi à identifier les différents phénomènes linguistiques à prendre compte pour le repérage automatique des expressions définitoires dans les textes.

Le second intérêt de cet article est de proposer un repérage automatique d'informations de la plus haute importance pour la lexicologie et la description sémantique. Nous pensons également avoir démontré que ce qu'on appelle « définition » doit être redécomposé en informations sémantiques relationnelles plus ultimes, parmi lesquelles les relations « classiques » (de catégorisation, de spécification, de synonymie, d'antonymie, complémentaire, de méronymie/holonymie) occupent une large place mais où la langue laisse la place à des expressions que l'on ne peut autrement catégoriser que comme des propriétés distinctives. Le repérage automatique de telles informations serait d'une très grande utilité pour les linguistes et lexicographes, et une suite à ce projet doit permettre de récupérer les contextes définitoires dans des textes de linguistique.

Bibliographie

- Arnauld (A.), Nicole (P.), 1662 : *La Logique ou l'Art de penser*, rééd. 1970, Flammarion.
- Borillo (A.), 1985 : « L'analyse automatique de textes : repérage et identification de la relation lexicale d'hyponymie », dans *Colloque Lexique, syntaxe et analyse automatique des textes*, Paris-X Nanterre, 3-4-5 mai 1995.
- Cartier (E.), 2004 : *Repérage automatique des expressions définitoires : modélisation de l'information définitoire, méthode d'exploration contextuelle, méthodologie de développement des ressources linguistiques, description des expressions du français contemporain, implémentation informatique*, Thèse de doctorat de l'Université Paris-IV Sorbonne.
- Cartier (E.), 2007 : « TextBox, a Written Corpus Tool for Linguistic Analysis », dans Fairon (C.), Naets (H.), Kilgarriff (A.), De Schryver (G.-M.), éd., *Building and Exploring Web Corpora (WAC3 - 2007)*, dans *Cahiers du CENTAL*, n° 4, Presses universitaires de Louvain, Louvain-la-Neuve, p. 33-42.
- Desclés (J.-P.), 1987 : « Réseaux sémantiques : la nature logique et sémantique des relateurs », dans *Langages*, n° 87, p. 55-78.
- Encyclopédie Philosophique Universelle, II — Les Notions philosophiques*, article *Définition*, Paris PUF, 1990.

- Granger (G. G.), 1976 : *La théorie aristotélicienne de la science*, Paris, Aubier-Montaigne.
- Kleiber (G.), 1984 : « Dénominations et relations dénominatives », dans *Langages*, n° 76.
- Kleiber, (G.), Tamba (I.), 1990 : « L'hyponymie revisitée : inclusion et hiérarchie », dans *Langages*, n° 98, p. 7-32.
- Chaurand (J.), Mazière (F.), éd., 1990 : « La Définition », dans *Actes du colloque La Définition*, organisé par le CELEX de l'Université Paris-Nord les 18 et 19 novembre 1988, Larousse.
- Langue française*, n° 73, février 1987, « La reformulation du sens dans le discours ».
- Martin (R.), 1976 : *Inférence, antonymie et référence*, Paris, Klincksieck.
- Martin (R.), 1983 : *Pour une logique du sens*, Paris, PUF.
- Rey (A.), 1965 : « À propos de la définition lexicographique », dans *Cahiers de lexicologie*, n° 6/1, Didier/Larousse, Paris.
- Rey-Debove (J.), 1978 : *Le métalangage*, Paris, Le Robert.
- Riegel (M.), 1987 : « Définition directe et définition indirecte dans le langage ordinaire : les énoncés définitoires copulatifs », dans *Langue française*, n° 73.

Hélène Ledouble,
Laboratoire Babel,
Université du Sud Toulon-Var,
Toulon,
France,
ledouble@univ-tln.fr.

Renate Trurnit Verbic,
Laboratoire CIRCPLES,
Université de Nice Sophia Antipolis,
Nice,
France,
jr.verbic@orange.fr.

Le « Vert » : couleur, signal, symbole. Étude lexicographique à partir d'un corpus quadrilingue

Résumé

Dans le cadre d'une recherche en contexte plurilingue, nous proposons une réflexion sur l'emploi du lexème *vert* et ses différentes combinaisons lexicales au sein de quatre langues européennes issues de trois groupes linguistiques distincts : langues latines (français), langues germaniques (anglais, allemand) et langues slaves (slovène). Notre objectif est de recenser et de classer les combinaisons lexicales contenant le lexème *vert* (*green/grün/zelen*) dans son contexte d'utilisation. Le plan de classement qui en résulte permet d'apparier ces combinaisons par langue et catégorie d'emploi (couleur, signal, symbole) dans une perspective de traduction et d'acquisition de connaissances à partir de corpus.

Mots-clés : corpus multilingue ; combinaisons lexicales ; traduction ; lexème *vert/green/grün/zelen*.

Introduction

Sous le titre *Le « Vert » : couleur, signal, symbole – Étude lexicographique à partir d'un corpus quadrilingue*, nous proposons une étude sur l'emploi du lexème *vert* et de ses équivalents au sein de quatre langues européennes, appartenant à trois groupes linguistiques différents : le français (langues latines), l'anglais, l'allemand (langues germaniques) et le slovène (langues slaves)¹.

¹ Le slovène est intéressant, n'ayant acquis que récemment le statut de langue nationale d'un nouvel État européen, créé en 1991.

Cette communication présente le travail de recensement et de classement de combinaisons lexicales contenant ces lexèmes et attestées dans un corpus journalistique en quatre langues.

Trois thématiques constituent la structure de notre classement: le vert employé en tant que couleur, signal ou symbole. Une mise en correspondance des quatre corpus, pour chacune de ces thématiques, nous permet d'identifier et de comparer les combinaisons lexicales équivalentes dans une perspective de traduction « basée sur les corpus » (Isabelle et Warwick-Armstrong 1993 : 288).

Notre communication s'articule en deux parties: nous présentons tout d'abord le corpus quadrilingue, le choix des sources utilisées, sa taille et ses principales caractéristiques lexicométriques. La seconde partie s'intéresse au plan de classement des combinaisons lexicales en fonction des trois thématiques. Nous concluons en situant notre travail de recherche par rapport à la thématique du colloque et en proposant des pistes de réflexion sur la continuité de ce projet.

1 Présentation du corpus

Nous avons choisi de travailler sur quatre langues européennes le français, l'anglais, l'allemand et le slovène. Ces quatre langues n'ont pas la même importance dans la communication internationale: les 2 millions de locuteurs slovènes² ne peuvent concurrencer les 330 millions de locuteurs anglais (en 1^{re} langue)³, l'allemand étant parlé par environ 90 millions⁴ de locuteurs et le français par 75 millions⁵ de locuteurs natifs.

1.1 Composition du corpus

Notre corpus quadrilingue est constitué principalement d'articles issus de grands journaux nationaux en ligne: les quotidiens *Le Monde* (français) et *The Times* (anglais), *Die Zeit* et *Süddeutsche Zeitung* respectivement pour l'Allemagne du Nord et du Sud, ainsi que *Handelsblatt*⁶ (finances, économie). Enfin, le quotidien *Delo* a été utilisé pour les textes slovènes. Les archives des dix dernières années ont été consultées: la recherche d'articles contenant le lexème *vert* et ses équivalents (*green/grün/zelen*) dans les différents journaux a permis d'obtenir quatre ensembles de textes journalistiques monolingues. De ces documents nous avons extrait les combinaisons lexicales composées du lexème *vert* (*green/grün/zelen*)

2 Population slovène au 30 juin 2006: 2 008 516 habitants. Sources: *Statistični urad Republike Slovenije, Ministrstvo za notranje zadeve — Centralni register prebivalstva, Ministrstvo za notranje zadeve — Direktorat za upravne notranje zadeve, Inštitut za varovanje zdravja.*

3 Les chiffres cités par Malherbe (1990 :22). sont proches de ceux de l'OCDE de 2008 pour l'Allemagne et de 2006 pour la France.

4 Source: *Statistisches Bundesamt Deutschland et Statista 2009 Quelle: OECD.* Nombre d'habitants en Allemagne en 2008: environ 82 millions, locuteurs allemands en Suisse et Autriche non compris.

5 L'OCDE donne pour la France 63 195 000 habitants en 2006.

6 La collecte à partir de plusieurs journaux allemands est justifiée par la le système fédéral de l'Allemagne contrairement à la France, à l'Angleterre et à la Slovénie.

pour obtenir quatre corpus de combinaisons lexicales monolingues⁷. Le nombre d'occurrences de chaque lexème est respectivement de : 1 200 (*vert*), 1 100 (*green*), 750 (*grün*) et 967 (*zelen*).

1.2 Corpus et thématiques de classement

La langue française jouant le rôle de pivot, nous avons traduit l'ensemble des collocations en français pour faciliter l'appariement thématique des quatre corpus. Chaque corpus de combinaisons lexicales a été classé selon notre plan de classement : le vert employé en tant que couleur, signal et symbole, en se basant sur les définitions suivantes⁸ :

- Couleur: Qualité de la lumière que renvoie un objet et qui permet à l'œil de le distinguer des autres objets, indépendamment de sa nature et de sa forme ;
- Signal: Signe convenu par lequel quelqu'un donne une information, un avertissement ;
- Symbole: Élément textuel dont la signification concrète est liée par une correspondance analogique à une signification abstraite qu'il évoque ou représente.

Dans le corpus quadrilingue, les combinaisons se répartissent selon les trois axes comme indiqué dans le tableau 1, ci-dessous :

| Langue | Couleur | Signal | Symbole |
|--------------|---------|--------|---------|
| Français | 149 | 388 | 72 |
| Anglais | 130 | 444 | 41 |
| Allemand | 148 | 464 | 129 |
| Slovène | 109 | 248 | 35 |
| Nombre total | 536 | 1544 | 277 |

Tableau 1

Le vert en tant que signal est le thème qui regroupe le maximum de combinaisons. La différence entre le nombre total du classement et le nombre des combinaisons extraites est principalement due aux répétitions de certaines combinaisons.

1.3 Caractéristiques lexicométriques

Les principales caractéristiques lexicométriques ont été obtenues à l'aide de l'outil de statistique textuelle *Lexico3* (Zimina 2005 : 117) et sont représentées par le tableau 2 page suivante.

7 L'extraction des collocations et l'étude lexicométrique ont été réalisées avec le logiciel de lexicométrie *Lexico 3*, mis à notre disposition par l'équipe universitaire SYLED-cl2AT, de l'Université Paris 3.

8 Définitions issues du TLFi, *Trésor de la langue française informatisé*.

| | Corpus français | Corpus anglais | Corpus allemand | Corpus slovène |
|----------------------|-----------------|----------------|-----------------|----------------|
| Nombre d'occurrences | 57310 | 99 440 | 31 824 | 42 177 |
| Nombre de formes | 10616 | 13 779 | 9 230 | 12 990 |
| Fréquence maximale | 2 792 | 4 737 | 1 153 | 1 461 |
| Nombre d'hapax | 6 371 | 7 179 | 6 171 | 8 104 |

Tableau 2

Bien que les corpus de textes français et anglais soient plus importants que les corpus allemand et slovène, le nombre de formes de l'ensemble des corpus est relativement proche. Cela s'explique par la nature flexionnelle de l'allemand et du slovène, qui entraîne un nombre de formes plus élevé. Ceci est aussi à l'origine du nombre important d'hapax dans ces deux langues.

L'étude des segments répétés dans les quatre corpus montre une spécificité propre à chaque langue dans l'emploi du vert : en français et en allemand par exemple, la couleur renvoie principalement au parti politique *Les Verts*. En additionnant les occurrences, le français compte 433 segments et l'allemand 215 segments faisant référence aux Verts. Dans le corpus anglais, le parti *The Greens* n'est représenté que par quelques occurrences (21) de la même manière que *les Verts de Slovénie* (36).

Suivent d'autres thèmes selon la langue : les *espaces verts* (11) en français, l'expression anglaise *green shoots of recovery*⁹ (20), la *couleur verte* (56) et la *génétique verte* (24) pour l'allemand. Le slovène mentionne fréquemment *zelena knjiga* (20), le « Livre vert » (sur la libre circulation des travailleurs en Europe), et *zelena barva* (la couleur verte) (15). La couleur verte est absente des segments répétés (supérieurs à 10) des corpus français et anglais.

2 Analyse du corpus et plan de classement

Les segments collectés dans les corpus constituent des combinaisons lexicales qui relèvent de plusieurs appellations en fonction de leur composition : il peut s'agir de collocations au sens donné par Firth¹⁰ (1957) ou de combinaisons lexicales plus libres. L'objet de cet article n'étant pas une analyse linguistique de ces segments, mais leur classement tripartite, nous gardons ici le terme générique de *combinaisons lexicales*, déjà utilisé par L'Homme (1997 : 15).

Nous avons opéré un classement thématique des différentes combinaisons lexicales dont nous donnons le détail en annexe. Quelques aspects de ce classement sont présentés dans le cadre de cet article.

9 Cette expression fait référence à des « signes de croissance » dans le domaine de l'économie, avec un jeu de mots récurrent sur des signes de croissance « verts », c'est-à-dire liés au respect de l'environnement.

10 « Les combinaisons lexicales qui relèvent d'une affinité de par leur sémantique, responsable de l'apparition simultanée et fréquente de deux mots dans une unité syntaxique »

2.1 Vert en tant que couleur

Cette thématique reprend l'emploi du vert avec son sens initial de *couleur* et de ses nuances.

2.1.1 Nuances

Les nuances du vert s'expriment par l'ajout d'un adjectif au lexème *vert*, exprimant l'intensité plus ou moins forte de la couleur. Ces adjectifs composés sont présents dans chaque langue pour nuancer la couleur, allant de l'usage le plus classique de la couleur (*vert pâle*, *vert clair*, *vert foncé*, etc.) à un usage spécifique à une langue. Les paires d'oppositions se retrouvent généralement dans l'ensemble des corpus, à l'image de :

[1] pâle/foncé (fr.), *pale/dark* (ang.), *hell/dunkel* (all.), *svetlo/temno* (slo.).

Le français mentionne le :

[2] *vert criard*, dont l'équivalent allemand le plus proche serait *grellgrün*, *intensivgrün* (vert vif/intense) et qui est comparable à l'anglais *bright green*.

L'anglais emploie :

[3] *lime green* ou *mottled-green*

qui ne présente pas d'équivalents dans nos autres corpus.

Le français décrit le vert du dollar par un *vert acide* et fait de manière générale une subtile distinction entre le *vert foncé* (intensité de la couleur) et le *vert sombre* (intensité de l'éclairage), distinction non retrouvée dans les autres corpus.

2.1.2 Comparaison avec le monde végétal et minéral

Les nuances de la couleur verte sont également rendues par la comparaison avec des éléments présents dans la nature. Les végétaux offrent une large palette de couleurs pour décrire les différents verts. Ci-dessous quelques exemples issus d'une longue liste :

[4] L'herbe (*vert gazon/grasgrün/travnato zelen*).

[5] L'épinard (*vert épinard/spinatgrün*).

[6] La pistache (*pistaziengrün*).

[7] L'asperge (*asparagus-like greenness*).

[8] L'olive (*olive green/olivengrün/olivno zelen*).

Il est à noter que les combinaisons lexicales françaises composées de deux mots (ex. : *vert gazon*, *vert pomme*), deviennent généralement une seule et même unité en allemand (*grasgrün*, *apfelgrün*).

La comparaison au vert des arbres est différente en français et en slovène. Le français connaît le

[9] *vert tilleul*.

Le slovène fait référence au « vert hêtre » avec

[10] *bukova zelen*.

Les minéraux servent également de référence dans les quatre langues :

[11] L'émeraude (*vert émeraude/emerald green/smaragdgrün/smaragdno zelen*).

[12] Le jade (*vert jade/green as jade/jadegrün/žadasto zelen*).

Enfin, la langue française fait aussi référence à son histoire et à sa culture pour la description de la couleur verte. Ainsi, nous retrouvons le *vert absinthe*, le *vert empire* et le *vert céladon*¹¹ qui n'ont pas d'équivalents dans les autres langues.

2.1.3 *Signe distinctif*

Cette section illustre le rôle de la couleur en tant qu'élément différenciateur. L'adjectif permet par exemple une sous-catégorisation des espèces animales ou végétales par rapport au terme générique : ainsi, le chou vert sera différent du chou frisé, du chou-fleur, etc. Cette différenciation par la couleur s'applique également aux insectes, aux reptiles ou aux crustacés : l'araignée verte se distingue de l'araignée rouge, le mamba vert du mamba noir ou la crevette verte de la crevette grise.

Dans le monde militaire ou du cyclisme, la couleur de l'habillement identifie un corps d'armée ou un sportif vainqueur du jour du Tour de France : le béret vert s'oppose alors au béret blanc, le maillot vert au maillot jaune. L'utilisation de la couleur verte pour marquer une distinction significative fait la transition vers l'utilisation du vert en tant que signal.

2.2 Vert en tant que signal

Nous présentons dans cette partie trois domaines dans lesquels le vert est régulièrement employé en tant que signal : la signalétique routière, la politique ainsi que l'économie.

2.2.1 *Signalétique routière*

La signalétique routière utilise la couleur verte pour signaler une autorisation de passage (ou de démarrage). La voie libre pour les automobiles ou les piétons est indiquée par :

[13] *feu vert*, ou *green man*.¹²

¹¹ Vert céladon = vert pâle (*Petit Robert*).

¹² *Green horse* (à Buckingham palace).

La synchronisation des feux qui permet de rouler sans devoir s'arrêter est la

[14] *vague verte* (*green wave/grüne Welle/zeleni val*¹³).

Cette même idée est reprise pour signifier un accord pour une action donnée, avec une légère variante selon les langues. En français on donne :

[15] *le feu vert*.

Les trois autres langues donnant :

[16] la « lumière » verte (*green light/grünes Licht/zelena luč*).

En signalétique, le vert est généralement opposé au rouge (*feu rouge/feu vert*), mais l'expression inverse (*donner le feu rouge*) est absente du corpus.

2.2.2 Signal politique

Le vert signale aussi un engagement écologique. Les partis politiques se préoccupent des questions environnementales, ce qui se reflète dans les « tendances vertes » en politique :

[17] *Engagement vert, élections vertes, idéologies vertes, capitalisme vert, etc.*

Les programmes politiques s'habillent de vert :

[18] *New Deal vert* (*Green New Deal/ökologisches New Deal/zeleni New Deal*).

[19] *Livre vert* (*Green Paper/Grünbuch/zelena knjiga*).

[20] *Ceinture verte* (*the Green Belt/Der Europäische Grüne Gürtel et Evropska zelena vez* (le lien/la jonction vert(e)).

Ces programmes européens sont particulièrement présents dans le corpus slovène, à la suite de la récente intégration de la Slovénie à l'Union européenne, ainsi que d'autres références à la politique européenne (*politique verte européenne, parti vert européen, EGP, European Green Institute, etc.*)

Avec les récents succès des Verts aux élections, ces derniers sont devenus la cible des journalistes, qui jouent avec le vert, en l'associant à des mots évocateurs d'événements politiques significatifs :

[21] Les Khverts/les Khmers verts¹⁴ (*vs Khmers rouges du Cambodge*).

[22] Dany le Vert (*vs Dany le Rouge, alias Daniel Cohn-Bendit en 1968*).

[23] Le péril rouge-vert (*vs Le péril jaune/le péril rouge*).

13 À noter, une traduction française pour ce terme, utilisée au Québec : *progressive signal system*.

14 Maurice Tournier, dans son article « Petite suite en vert », citait déjà cette expression issue du *Nouvel Observateur*, 11 juin 1992 : « La chasse aux Khmers verts » (lire « La chasse aux écologistes intégristes »).

[24] Les Verts parlent aux Verts (*vs* « Les Français parlent aux Français » en 1940¹⁵).

La presse allemande parle des :

[25] *grauen Grünen* (les Verts « gris », signifiant « grisonnant »),

fustigeant ainsi le non-renouvellement des générations chez les Verts allemands.

Lors des récentes manifestations d'étudiants en Iran, la presse allemande parlait d'un :

[26] « tsunami vert » (*der grüne Tsunami*),

signifiant qu'il s'agissait là d'un raz de marée destructeur pour le gouvernement en place.

2.2.3 Signal économique

Dans le domaine de l'économie et des finances, une évolution des mentalités se reflète dans les différents programmes (tableau 3) :

| Thèmes | FR | ANG | ALL | SLO |
|--------------------|---|---|---|---|
| Économie générale | <i>croissance verte</i> <i>économie verte</i> <i>PIB vert</i> | <i>green growth</i> | <i>sie wollten nicht nur die grünen Wirtschaftsbereiche fördern</i> | <i>zelena rast</i> <i>Naložbe v zeleno ekonomijo</i> |
| Entreprise emploi | <i>emplois verts</i> <i>géant vert</i> <i>secteur vert</i> | <i>green collar jobs</i> <i>green fortunes,</i> <i>green entrepreneur</i> | | |
| Fiscalité bancaire | <i>taxes vertes,</i> <i>prêt à taux zéro vert</i> <i>obligations vertes</i> <i>livret vert</i> | <i>green taxation</i> <i>green fiscal stimuli</i> <i>green loans</i> | <i>damit auch finanziell alles im grünen Bereich bleibt.</i> | |

Tableau 3

15 Après le célèbre « Appel du 18 juin 1940 », s'adressant aux soldats et à leurs officiers, le général de Gaulle lance, le 14 juillet 1940, une émission quotidienne en français sur les ondes de la BBC, à Londres : « Les Français parlent aux Français ».

À l'image des campagnes de Greenpeace, dans les années 1980, la tendance touche toutes les activités, y compris l'urbanisme et la construction (tableau 4) :

| Thèmes | FR | ANG | ALL | SLO |
|--------------|--|-----------------------|---|--|
| Urbanisme | <i>capitales vertes</i> <i>villes vertes</i> | <i>greener London</i> | <i>grünes Ratingen</i> | <i>Zelenojamsko</i> « <i>zamestje</i> » |
| Construction | <i>toitures vertes</i> <i>maison verte</i> <i>béton vert</i> | <i>green roofs</i> | <i>Gründdach</i> <i>grünes</i> <i>Baumaterial</i> <i>grüne Gebäude</i> | <i>Hiša z zeleno</i> <i>streho</i> <i>koncept zelene</i> <i>gradnje</i> |

Tableau 4

Dans le secteur des biens de consommation, cette propension est également très nette :

[27] *détergents verts, ordinateurs verts, chaussures vertes, piles vertes, corbillard vert*, etc.

Cette déferlante répond à une double exigence :

- de nouvelles contraintes sociétales et gouvernementales incitent les acteurs économiques à un comportement plus responsable dans la production et la vente de produits et services ;
- elle répond également à l'attente des consommateurs désireux de se donner bonne conscience dans leur vie quotidienne et leurs achats.

Après le « tourisme vert », la tendance verte s'étend maintenant à l'ensemble du secteur touristique et des loisirs :

[28] des *restaurants verts* ;

[29] *green drinking* (« boire vert ») ;

[30] la « lune de miel verte » (ang. : *green honeymoon*).

La publicité affiche la couleur pour signaler un « engagement écologique », réel ou souhaité (les vertes affiches, labels et autres logos verts).

L'utilisation du vert en communication active certains automatismes « favorables », mais reste souvent ambiguë. En effet, certaines combinaisons lexicales restent équivoques hors contexte, car les mêmes combinaisons (*restaurant vert, ligne verte*, etc.) ont, au sein d'une même langue, des significations différentes :

- le « restaurant vert » par exemple, fait référence soit à la nourriture à base de légumes, soit à un établissement construit à partir de matériaux respectant l'environnement ;

- la « ligne verte » peut désigner plus de trois éléments distincts au sein même du corpus français (la ligne de cessez-le-feu entre Israël et la Cisjordanie, le futur métro de Los Angeles, une ligne séparant la zone turque de la zone grecque (également appelée *zone Attila*), etc.).

Les différences sont également importantes entre les langues. Ainsi, l'« autoroute verte » française permet aux Parisiens de rejoindre rapidement la verdure du Massif central, alors que la *Green Highway* (littéralement « autoroute verte ») de Californie est une autoroute au bord de laquelle l'on ne trouve que des stations de carburants « verts » au sens de « propres ».

L'emploi du « vert », déjà important au début des années 1990 (Tournier 1994), ne cesse de s'intensifier. Utilisé en tant que signal, ce lexème peut ainsi revêtir de nombreuses significations, comme l'attestent quelques exemples issus du corpus :

- l'absence de produits dopants (*lacet vert*) ;
- l'engagement écologique (*flamme verte*) ;
- les matériaux non polluants (*béton vert*) ;
- les matières recyclables (*pile verte*) ;
- la gratuité (*numéro vert*) ;
- la santé (*croix verte*).

Comme nous l'avons montré, l'information dépasse le sens littéral des combinaisons lexicales. À la nécessité d'informer s'ajoute le besoin de convaincre lecteurs et consommateurs. Dans ce contexte, on constate que le lexème *vert* en tant que signal est un puissant outil de stratégie de communication politique et commerciale. Ce signal peut lui-même devenir symbole, évoquant une réalité pour une culture donnée, dont la compréhension du sens est réservée aux initiés. Nous présentons ici la troisième thématique de ce classement, l'emploi du vert en tant que symbole.

2.3 Vert en tant que symbole

Sous ce thème, nous trouvons les segments liés aux religions, aux cultes, aux valeurs patriotiques, aux valeurs symboliques positives ou négatives du vert.

Cette symbolique est exprimée explicitement :

[30] *zelena je svetovna islamska barva* (littéralement : « Le vert est la couleur sacrée de l'islam »).

ou implicitement :

[31] *Grün erscheint als Einbandfarbe von islamischen Büchern* (littéralement : « Le vert est utilisé pour relier des livres islamiques »).

En [31], la symbolique du sacré est uniquement suggérée au lecteur par l'association de la couleur verte avec l'islam.

2.3.1 Valeurs symboliques culturelles de l'islam et de la chrétienté

La couleur verte sert de symbole aux religions musulmane et catholique :

- couleur sacrée, couleur du paradis et couleur d'immortalité pour l'islam
 - [32] *Grün ist die Kultfarbe des Islam* (littéralement: « C'est pour cette raison que le vert est la couleur culturelle de l'islam »).
 - [33] *zelena je svetovna islamska barva* (littéralement: « Le vert est la couleur sacrée de l'islam »).
 - [34] un caveau drapé d'un tissu vert (islam).
- couleur liturgique pour l'Église catholique :
 - [35] *Liturgična barva med navadnim letom je zelena* (littéralement: « La couleur liturgique pendant l'année ordinaire est verte »).

Le vert saturé avait une valeur positive et symbolisait la couleur de la Trinité, de l'Épiphanie et de saint Jean l'Évangéliste (Probiesch 2006).

2.3.2 Du culturel au politique

La valeur symbolique culturelle du vert glisse vers le politique lorsque la couleur est présentée sur les drapeaux nationaux. Le vert recouvre la totalité des drapeaux de l'Arabie Saoudite, de la Libye et de la Mauritanie. Sur les drapeaux d'autres pays du désert, le vert peut être autant symbole de religion que symbole de survie, car représentant l'oasis, assurant la survie dans un milieu hostile :

- [36] *Grün findet sich auch in den Nationalflaggen der meisten Wüstenstaaten, das Überleben in den Weiten der endlosen Wüsten ist nur gesichert wenn man rechtzeitig eine grüne Oase erreicht* (littéralement: « Nous trouvons le vert dans les drapeaux de nombreux États du désert, parce que la vie dans les déserts sans fin est garantie seulement si on arrive à temps dans une oasis verte »).

La nuance de la couleur semble avoir son importance :

- [37] *Islamske politične stranke, imajo najraje zeleno barvo v enakem odtenku* (littéralement: « Les partis politiques islamiques préfèrent la couleur verte dans une même nuance »).

Le vert est aussi la couleur des partis islamiques :

- [38] *the green flags of Mr Sharif's Pakistan Muslim League* (littéralement: « Les drapeaux verts de la Ligue musulmane du Pakistan de M. Sharif »).

La puissance de la symbolique verte a également été démontrée par la *marche verte* [au Sahara, 1975] pour l'unité et l'indépendance du Sahara marocain.

Les journalistes jouent de l'identification de l'islam avec la couleur verte, comme dans l'exemple suivant :

[39] Le péril vert,

s'oppose ici au « péril rouge », formule longtemps utilisée pour fustiger le danger communiste et qui suggère de ce fait le danger de l'islam politique.

Cependant, le vert est aussi présent sur les drapeaux de pays non islamiques, symbolisant des valeurs fortes, mais différentes selon les pays :

– Irlande: religion catholique et couleur nationale

[40] *In Irland, [...], wird die Farbe Grün als Nationalfarbe von den Katholiken beansprucht.* (littéralement: « En Irlande, la couleur verte est revendiquée comme couleur nationale par les catholiques »).

– Brésil: la maison de Bragança (l'Empereur brésilien Pedro I).

[41] *Grün ist vielmehr die Farbe des Hauses Braganza* (littéralement: « Le vert est plutôt la couleur de la maison de Bragança »).

– Italie: le droit des hommes à la liberté et à l'égalité:

[42] *In der grün-weiß-roten Flagge Italiens [...] symbolisiert das Grün das Recht des Menschen auf Freiheit und Gleichheit.* (littéralement: « Sur le drapeau italien vert-blanc-rouge, le vert symbolise le droit de l'homme à la liberté et à l'égalité »).

Le vert symbolise donc selon les pays une valeur forte ancrée dans la culture.

2.3.3 Autres valeurs symboliques positives ou négatives

Sur un plan plus large, le vert représente l'espoir, la renaissance et la fécondité en rappelant le retour annuel du printemps :

[43] *Zelena barva je barva življenja in rastlin* (littéralement: « la couleur verte est la couleur de la vie et des plantes »).

La couleur verte peut aussi symboliser un manque de maturité, l'inexpérience, la toxicité et le mal :

[44] *giftgrün* (vert poison).

[45] *Zeleno povezuje tudi z nezrelostjo, strupenostjo in demonskostjo.* (littéralement: « Nous rattachons aussi le vert à un manque de maturité, à la toxicité et au diabolisme »).

[46] *Die Farbe Grün ist aber auch Symbol für Unerfahrenheit* (littéralement: « la couleur verte est aussi un symbole d'inexpérience »).

Cette symbolique négative est également suggérée dans le film *The Green Mile* dans lequel le sol du couloir qui mène les condamnés à mort à la chaise électrique est recouvert de linoléum vert.

La notation négative de la couleur verte est aussi présente dans plusieurs locutions (tableau 5) :

| Français | Anglais | Allemand | Slovène |
|---|--|--|---|
| vert de jalousie, vert de rage vert de peur | <i>gadget that will turn your mates green with envy</i> (littéralement: « devenir vert d'envie/ de jalousie ») | <i>grün und gelb werden vor Ärger, Neid</i> (littéralement: « devenir vert et jaune de colère, de jalousie ») | <i>biti zelen od jeze, strahu, zavisti</i> (littéralement: « devenir vert de rage, de peur, de jalousie ») |

Tableau 5

Selon les langues, on peut être vert de jalousie, de rage, de peur ou de colère; en allemand, la connotation négative du vert est renforcée par la couleur jaune, évoquant l'idée de poison :

[47] *Es existiert bis heute die Vorstellung, dass Grün zusammen mit der Farbe Gelb mit dem Begriff „Gift“ verbunden ist* (littéralement: « Il existe jusqu'à aujourd'hui l'idée que le vert, associé au jaune, est lié au concept de poison »).

Les couleurs ne sont pas interchangeable, elles sont culturellement liées et activent dans chaque langue une image spécifique faite de références culturelles et d'émotions (Dobrovols'kij 1997 : 106)¹⁶ et sont, dans ce contexte, difficilement traduisibles.

Conclusions et perspectives

L'étude des emplois du lexème *vert* en quatre langues était appropriée pour explorer la thématique du colloque « Passeurs de mots, passeurs d'espoir : lexicologie, terminologie et traduction, face au défi de la diversité ». L'extraction des combinaisons lexicales, leur classement et appariement thématique ont constitué un premier pas vers l'élaboration d'une ontologie du vert. Ce classement peut faciliter « l'accès à des ressources lexicales et phraséologiques bilingues » (Zinglé 1999 : 11), et ainsi servir dans le cadre d'applications d'aide à la traduction ou à la rédaction, ainsi

¹⁶ Littéralement: « La prise en compte de la motivation de la forme interne d'un idiome, pouvant être une image métaphorique, un symbole ou un quasi-symbole (cf. : 2.3.2), est importante pour découvrir les structures conceptuelles qui s'y trouvent. Ces structures sont liées culturellement et influencent la perspective de la réalité. Le locuteur, en utilisant un idiome, propose à l'interlocuteur de regarder ce fragment du monde d'une manière spéciale. La forme (en tant qu'image) de l'expression influence particulièrement les sphères qui sont inaccessibles à une observation directe. »

que pour l'enseignement de langues étrangères. En effet, au-delà de l'étude de la langue et de sa traduction, l'exploration d'un corpus multilingue dévoile le regard similaire ou distinct que portent les francophones, les anglophones, les germanophones ou les Slovènes sur le « phénomène vert ».

Nous avons vu que le vert, réputé couleur instable, est également instable dans le message qu'elle transmet. Selon le regard, elle peut représenter le bien ou le mal, l'espoir ou la malédiction. Actuellement, la couleur verte symbolise l'espoir d'un changement qui recouvre selon l'actualité une thématique différente: indépendance (Sahara marocain), liberté et changement politique (Iran), ou changement d'habitudes et de mentalités dans nos pays fortement industrialisés (contre le gaspillage des ressources, pour un commerce plus équitable, une meilleure protection de l'environnement, etc.). Dans ce contexte, partis politiques et industriels utilisent la couleur verte tel un code d'honneur, garantissant aux électeurs et consommateurs des pays développés un comportement pleinement en accord avec leurs exigences actuelles.

Bibliographie

- Bajec (A.) *et al.*, 1998 : *Slovar slovenskega knjižnega jezika*, éd. Slovenska akademija znanosti in umetnosti [in] Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti, Inštitut za slovenski jezik Frana Ramovša.- Ljubljana : DZS (Slovarji DZS).
- Dobrovols'kij (D.), 1997 : *Idiome im mentalen Lexikon: Ziele und Methoden der kognitivbasierten Phraseologieforschung*, Trier, WVT Wissenschaftlicher Verlag (Fokus Linguistisch-Philologische Studien, n° 18).
- Firth, (J.R.), 1957 : « Papers in Linguistics 1934-1951, Chapter *Modes of meaning* », éd., Oxford University Press, London, p. 190-215. (Publié d'abord in *Essays and Studies*, the English association, 1951.)
- Herrity (P.), 2000 : *Slovene: a comprehensive grammar*, London, Routledge, (Routledge Grammars).
- Isabelle (P.) et Warwick-Armstrong (S.), 1993 : « Les corpus bilingues : une nouvelle ressource pour le traducteur », éd. *La Traductique*, études et recherches de traduction par ordinateur, sous la direction de Bouillon P. et Clas A., Montréal, Presses de l'université de Montréal/AUPELF/UREF, p. 288-306.
- L'Homme (M.-C.), 1997 : « Méthode d'accès informatisé aux combinaisons lexicales en langue technique », dans *Meta*, vol. 42 n° 1, mars 1997, p. 15-23.
- Malherbe (M.), 1990 : *Les Langages de l'Humanité – Une encyclopédie des 3 000 langues parlées du monde*, Paris, Robert Lafont, (Bouquins).
- Probiesch (K.), 2006 : *Religiöse Feste, Feiertage und Gedenktage*, Marburg, dernière mise à jour le 15 juin 2006, www.feste-der-religionen.de/farben/gruen.
- Rey-Debove (J.), 1998 : *La linguistique du signe, Une approche sémiotique du langage*, Paris, Armand Colin (Linguistique).
- TLFi, 2009 : *Le Trésor de la Langue française informatisé*, 2009, France, Laboratoire ATILF. www.atilf.fr/tlfi.htm

- Tournier (M.), 1994 : « Petite suite en vert » dans *Mots*, juin 1994, n° 39 Environnement. Écologie. Verts. p. 114-116. doi : 10.3406/mots.1994.1895. url : http://www.persee.fr/web/revues/home/prescript/article/mots_0243-6450_1994_num_39_1_1895
- Zimina-Poirot (M.), 2005 : « Topographie bitextuelle et approches quantitatives de l'extraction de ressources traductionnelles à partir de corpus parallèles », dans Blampain (D.), Thoiron (Ph.) et Van Campenhoudt (M.), éd. *Mots, Termes et Contextes, Actes des VII^{es} Journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction*, (Haute École de Bruxelles – ISTI, 8,9 et 10 septembre 2005, Bruxelles, p. 175-186), (Universités francophones – actualité scientifique).
- Zinglé (H.), 1999 : « La modélisation des langues naturelles: Aspects théoriques et pratiques », *Travaux du LILLA* numéro spécial, Faculté des Lettres, Arts et Sciences humaines de l'université de Nice-Sophia Antipolis, Nice.

Annexe

Plan de classement thématique

- 1 Vert en tant que Couleur
 - 1.1 Couleur verte
 - 1.1.1 Dégradés du vert
 - 1.1.2 Nuances du vert par comparaison avec la nature
 - 1.2 Production de la couleur verte
 - 1.2.1 Pigments verts
 - 1.2.2 Lumière verte
 - 1.3 Usage du vert
 - 1.3.1 Usage du vert dans le monde végétal
 - 1.3.2 Usage du vert dans le monde animal
 - 1.3.3 Usage du vert dans le monde minéral
 - 1.3.4 Usage du vert dans la vie quotidienne
 - 1.3.5 Usage du vert dans le sport ou les jeux
 - 1.3.6 Usage du vert pour des personnages
 - 1.4 Relation de l'homme avec la couleur verte
 - 1.4.1 Attirance pour ou rejet de la couleur verte
 - 1.4.2 Effet révélateur d'un caractère
 - 1.4.3 Effet bénéfique ou thérapeutique
 - 1.4.4 Effet défavorable
 - 1.4.5 Couleur verte et art
 - 1.5 Vert (couleur) en tant que signe distinctif
 - 1.5.1 Agriculture
 - 1.5.2 Végétaux
 - 1.5.3 Fruits et Légumes
 - 1.5.4 Arbres

- 1.5.5 Epices
- 1.5.6 Thé
- 1.5.7 Insectes, reptiles et autres
- 1.5.8 Sport – Loisirs
- 1.5.9 Santé
- 1.5.10 Territorial
- 1.5.11 Économique
- 1.5.12 Divers

- 2 Vert en tant que signal
 - 2.1 Signalétique routière
 - 2.1.1 Feux de signalisation
 - 2.1.2 Assurance voiture internationale
 - 2.1.3 Expressions à partir de la signalétique routière
 - 2.2 Signalétique de solidarité – d'appartenance
 - 2.2.1 Appartenance à un club sportif
 - 2.2.2 Appartenance à une nation
 - 2.3 Signal d'engagement écologique (national)
 - 2.3.1 Tendances et idées vertes
 - 2.3.2 Engagement politique
 - 2.3.3 Ralliement à un parti/une cause
 - 2.3.4 Le vert en politique
 - 2.3.5 Les Verts et les enjeux électoraux
 - 2.3.6 Regards sur les Verts
 - 2.3.7 Programmes/agenda vert(s)
 - 2.4 Signal d'engagement écologique européen
 - 2.4.1 Qualité de la politique européenne
 - 2.4.2 Buts de la politique européenne
 - 2.4.3 Les partis verts européens
 - 2.4.4 Programmes de la politique européenne
 - 2.5 Signal d'engagement écologique international
 - 2.5.1 Discours Nord-Sud
 - 2.5.2 Discours Est-Ouest
 - 2.6 Signal d'une nouvelle économie
 - 2.6.1 Économie générale
 - 2.6.2 Emplois verts
 - 2.6.3 Finance verte
 - 2.6.4 Investissements verts
 - 2.6.5 Secteur vert
 - 2.6.6 Mesures/initiatives vertes
 - 2.6.7 Entreprises — *Business*

- 2.6.8 Bancaire
- 2.6.9 Énergie – Électricité
- 2.6.10 Technologie – Électronique
- 2.6.11 Produits verts – Consommation verte
- 2.6.12 Labels – étiquetage
- 2.6.13 Automobile – Voitures vertes

- 2.7 Signal d'un nouvel urbanisme
 - 2.7.1 Villes vertes
 - 2.7.2 Villes et surfaces vertes
 - 2.7.3 Construction verte
 - 2.7.4 Habitation
 - 2.7.5 Voies vertes – espaces verts
 - 2.7.6 Région, pays, continent, planète vert(e)

- 2.8 Signal d'une évolution d'une mentalité
 - 2.8.1 Voyages/Tourisme
 - 2.8.2 Loisirs
 - 2.8.3 Restauration
 - 2.8.4 Sport
 - 2.8.5 Évènements
 - 2.8.6 Tendances société – Mariage
 - 2.8.7 Mouvement/tendance/idéologie verte
 - 2.8.8 Le vert, valeur publicitaire

- 3 Vert en tant que symbole
 - 3.1 Symbolique générale
 - 3.2 Les valeurs symboliques culturelles du vert
 - 3.2.1 Couleur sacrée de l'islam
 - 3.2.2 Couleur verte et christianisme
 - 3.2.3 Couleur de la sorcellerie
 - 3.3 Du symbole culturel au symbole politique
 - 3.3.1 Couleur des drapeaux
 - 3.4 Culture différente, symbole différent
 - 3.4.1 Sens symbolique selon les pays
 - 3.5 Symbolique favorable/défavorable
 - 3.5.1 Autres valeurs symboliques favorables
 - 3.5.2 Symbolique négative
 - 3.5.3 Expression utilisant la symbolique du vert
 - 3.5.4 *Vert* entrant dans diverses expressions

Janine Pimentel,
Marie-Claude L'Homme,
Observatoire de linguistique Sens-Texte (OLST),
Université de Montréal,
Montréal,
Canada,
janine.pimentel@umontreal.ca.
mc.lhomme@umontreal.ca.

Annotation syntaxico-sémantique de contextes spécialisés : application à la terminographie bilingue

Résumé

Le présent article décrit un modèle d'annotation formelle des propriétés syntaxico-sémantiques des termes de nature prédicative et de leurs actants (également appelés *arguments*) à partir de contextes dans lesquels ces termes apparaissent. Nous nous pencherons sur les différentes adaptations qui ont été faites au modèle d'annotation utilisé au préalable pour les contextes contenant des verbes d'informatique français (ex. : *cliquer, compiler, configurer*) lorsque l'analyse de contextes contenant des verbes d'informatique anglais (ex. : *configure, access, click*) a débuté. Cette réflexion préalable s'appuie sur 50 verbes et sur environ 1 000 contextes annotés. Nous ferons également quelques observations préliminaires sur des différences de comportement linguistique entre les verbes anglais annotés et leurs équivalents français que l'annotation a permis de révéler.

Mots-clés : annotation syntaxico-sémantique ; contextes spécialisés ; terminographie bilingue.

Introduction

Les dictionnaires spécialisés et les banques de terminologie proposent des descriptions de nature conceptuelle souvent très riches, mais fournissent très peu de renseignements sur les propriétés linguistiques des termes ou sur leur comportement en langue. Quelques exceptions, toutefois, ont commencé à faire leur apparition : des dictionnaires spécialisés conçus dans une perspective d'apprentissage (par exemple, le Binon *et al.* 2000 dans le domaine des affaires) ou des bases terminologiques incorporant des rubriques sur le fonctionnement syntaxique ou sur la combinatoire des termes (par exemple, Lépinette et Olmo Cazevielle 2007 ; DiCoInfo, L'Homme 2008).

Pour remplir partiellement cette lacune, nous avons proposé un modèle ainsi qu'une méthode d'annotation contextuelle permettant d'explicitier les propriétés

syntaxico-sémantiques des termes de nature prédicative (plus précisément les verbes) (Hadouche *et al.* 2009 ; L'Homme *et al.* 2009). L'application du modèle et de la méthodologie s'inspirant largement de ce qui a été fait dans le cadre du projet FrameNet (Ruppenhofer *et al.* 2006) a été réalisée, dans une première étape, sur des verbes français appartenant aux domaines de l'informatique (ex. : *configurer, cliquer, télécharger*) et de l'environnement (ex. : *réchauffer, stocker, piéger*)¹. Nous l'avons ensuite appliqué à l'analyse de verbes anglais (ex. : *configure, access, click*) et envisageons une extension à l'espagnol. Même si le modèle d'annotation n'avait pas été restreint à l'origine à une seule langue, son application à l'anglais a exigé d'inévitables ajustements.

La section 2 décrit dans ses grandes lignes le modèle d'annotation adopté ainsi que la méthodologie mise au point pour annoter les termes. Ensuite, nous nous pencherons sur les différentes adaptations qui ont été faites au modèle d'annotation lorsque nous avons commencé à prendre en compte des contextes contenant des verbes d'informatique anglais (Section 3). Cette réflexion préalable s'appuie sur 50 verbes (Annexe 1) et sur environ 1 000 contextes annotés. Nous ferons également quelques observations préliminaires sur des différences de comportement linguistique entre les verbes anglais annotés et leur équivalent français que l'annotation nous a permis de révéler (Section 4). Quelques perspectives sont évoquées en conclusion.

1 Annotation de propriétés syntaxico-sémantiques

Le modèle que nous proposons consiste à annoter formellement les propriétés syntaxico-sémantiques des termes de nature prédicative et de leurs actants (également appelés *arguments*) à partir de contextes dans lesquels ces termes apparaissent. Les éléments explicités dans les annotations sont les suivants (un exemple d'application est donné au tableau 1) :

- terme prédicatif faisant l'objet de l'annotation (*configurer* dans le tableau 1) ;
- participants et leur nature (actants ou circonstants) : les participants sont subdivisés en actants (qui contribuent au sens du terme) (*système, cartes*, dans le tableau 1) et en circonstants (*automatiquement* et *informations*, dans le tableau 1) (qui ne contribuent pas au sens du terme, mais qui apparaissent dans les contextes ; *Autres* dans le tableau) ;
- rôle sémantique du participant (Agent, Patient, Instrument, etc.) ;
- fonction syntaxique du participant (sujet, objet, complément, modificateur) ;
- groupe syntaxique du participant (SN, SP, SAdv, Prop).

¹ Il est intéressant de noter que le modèle FrameNet semble particulièrement bien adapté à la description des termes appartenant à des domaines spécialisés. Il est utilisé, notamment, dans le domaine de la biomédecine (Dolbey *et al.* 2006), du droit (Venturi *et al.* 2009) et du football (Schmidt 2009).

Le système peut, avec ces informations, CONFIGURER automatiquement les cartes d'extension PCI.

| CONFIGURER 1 | | |
|--------------|----------------------|-----------------|
| Actants | | |
| Agent | Sujet (SN) | système |
| Patient | Objet (SN) | carte |
| Circonstants | | |
| Manière | Modificateur (SAdv) | automatiquement |
| Moyen | Complément (SP-avec) | information |

Tableau 1 : Propriétés syntaxico-sémantiques du terme configurer et de ses participants telles que réalisées dans un contexte.

Les annotations sont articulées autour des rôles sémantiques des participants (actants et circonstants) d'un terme prédicatif. Le rôle sémantique est défini comme le lien partagé par le participant et le terme prédicatif (Agent, Patient, Instrument, Manière, Destination). Malgré le caractère subjectif qu'attribue parfois la littérature à la notion de « rôle sémantique », la représentation des participants par des étiquettes de rôles se révèle particulièrement utile pour rendre compte des participants partageant le même lien avec des unités prédicatives différentes (Fillmore 1968; FrameNet 2009; VerbNet 2009).

L'exemple [1] montre de quelle manière les actants de termes reliés à *configurer* sont représentés au moyen d'étiquettes de rôles. Bien que la position syntaxique des actants varie, leur rôle reste le même.

[1] CONFIGURER_{1a}: Agent ~ Patient (ex.: *L'utilisateur configure une imprimante*)

CONFIGURATION_{1,1}: ~ de Patient par Agent (ex.: *Configuration de l'imprimante par l'utilisateur*)

CONFIGURABLE₁: Patient ~ (ex.: *Un clavier configurable*)

RECONFIGURER₁: Agent ~ Patient (ex.: *Le programmeur reconfigure le noyau*)

Les étiquettes de rôles sémantiques que nous utilisons s'apparentent aux étiquettes utilisées pour représenter les éléments d'un Frame (FE) dans FrameNet (2009). Toutefois, elles s'en distinguent en ce sens que nous tentons de définir un nombre limité d'étiquettes qui s'appliqueront à l'ensemble des termes dans un domaine spécialisé (et non uniquement à l'intérieur d'un seul Frame) ².

² Afin de conserver un parallélisme entre le français et l'anglais, les rôles ont immédiatement été ajoutés au schéma gérant l'annotation des contextes français.

Les annotations sont réalisées dans une structure XML gérée par un schéma forçant les annotateurs à respecter la structure des balises³. Le schéma (sorte de grammaire gérant la structure des fichiers XML) permet également de formaliser les annotations pour qu'elles puissent être utilisées par des applications de traitement automatique des langues (TAL). La figure 1 montre de quelle manière une phrase contenant le terme *configurer* est annotée en XML.

```
< contexte source = "ASSDELL" statut = "o" annotateur = "MCLH" mise-a-jour = "2008-03-06">
  <contexte-texte> Le système peut... configurer automatiquement les cartes d'extension PCI.
</contexte-texte>
  <participant type = "Act" role = "Agent">
    <fonction-syntaxique nom = "Sujet">
      <groupe-syntaxique nom = "SN"> Le
        <realisation> système </realisation>
      </groupe-syntaxique>
    </fonction-syntaxique>
  </participant> peut...
  <lexie-att> configurer </lexie-att>
  <participant type = "Circ" role = "Manière">
    <fonction-syntaxique nom = "Modificateur">
      <groupe-syntaxique nom = "SAdv">
        <realisation> automatiquement </realisation>
      </groupe-syntaxique>
    </fonction-syntaxique>
  </participant>
  <participant type = "Act" role = "Patient">
    <fonction-syntaxique nom = "Objet">
      <groupe-syntaxique nom = "SN"> les <realisation lemme = "carte"> cartes </realisation>
        d'extension PCI </groupe-syntaxique>
    </fonction-syntaxique>
  </participant>.
</contexte>
```

Figure 1 : Annotation d'un contexte du terme configurer.

Pour chaque terme de nature prédicative, nous extrayons entre 15 et 20 contextes dans lesquels ces termes apparaissent. Soulignons également que les fiches sont consacrées à des acceptions et non à des formes linguistiques. Ainsi, les contextes extraits illustrent l'acception décrite. Les contextes sont prélevés dans des corpus spécialisés (partiellement comparables) : le corpus français comporte un peu de plus de 2 000 000 d'occurrences; le corpus anglais compte environ 1 000 000 d'occurrences. Les contextes servent à illustrer de quelle manière le terme s'utilise concrètement dans les textes spécialisés et ils sont choisis selon des critères de représentativité du corpus ainsi que de la variété du comportement linguistique des termes. Ces contextes sont par la suite annotés selon les paramètres signalés plus haut.

3 La structure XML, le schéma qui la gère ainsi que les scripts permettant de générer la représentation HTML présentée plus haut ont été réalisés en collaboration avec Fadila Hadouche et Guy Lapalme du Département d'informatique et de recherche opérationnelle de l'Université de Montréal.

2 Adaptation à l'anglais

L'annotation des contextes anglais dans l'éditeur XML oXygen suit, de manière générale, les mêmes étapes que l'annotation des contextes français (L'Homme *et al.* 2009). Cependant, l'annotation des contextes anglais a nécessité un travail d'adaptation de certaines parties du schéma français que nous détaillons ci-dessous.

2.1 Annotation des termes

La première étape du travail d'adaptation a porté sur la façon dont on annote les termes. Étant donné que les termes annotés sont des verbes, ceux-ci souvent peuvent être accompagnés de verbes auxiliaires ou dépendre syntaxiquement d'autres verbes. Puisque ces verbes qui accompagnent les termes changent d'une langue à l'autre, un traitement séparé a été défini pour l'anglais.

En français, les termes peuvent être accompagnés d'un des deux auxiliaires suivants : *être* ou *avoir*. Dans ces cas, nous annotons le terme ainsi que l'auxiliaire : ce dernier apparaît comme valeur de l'attribut *auxiliaire*. En anglais, outre les auxiliaires *be* et *have*, les termes peuvent être accompagnés de l'auxiliaire *do* (exemple 2). Une étape du travail d'adaptation a donc consisté à ajouter l'auxiliaire *do* au schéma d'annotation anglais.

[2] *How do I DOWNLOAD a file from the internet ?*

```
<lexie-att auxiliaire = "do"> do </lexie-att>...
```

```
<lexie-att> download </lexie-att>
```

Par ailleurs, en français les termes peuvent dépendre syntaxiquement d'autres verbes comme *aller*, *croire*, *devoir*, *faire*, *pouvoir*, *sembler*, *venir de* et *vouloir*. Dans ces cas, nous n'avons pas annoté ces verbes et nous avons considéré que les participants ont un lien syntaxique direct avec le terme. Par exemple, dans le contexte [3] le terme *taper* dépend syntaxiquement du verbe *venir de*. Pourtant, nous avons considéré que le participant *vous* est le sujet du terme *taper*.

[3] Le texte que vous venez de TAPER réside actuellement dans la mémoire vive...

De la même manière, les contextes anglais renferment des termes annotés qui dépendent syntaxiquement d'autres verbes. Par exemple, dans le contexte [4], le terme *activate* dépend syntaxiquement du verbe *should*. Pourtant, nous avons considéré que le participant *CD-ROM* est le sujet du terme.

[4] *The CD-ROM should ACTIVATE.*

Une étape du travail d'adaptation a donc consisté à identifier les verbes anglais qui suivent ce patron identifié au préalable dans l'annotation des contextes français. Ces verbes sont les suivants : *can*, *may*, *might*, *must*, *should*, *will* et *would*.

Dans tous les autres cas (français : *avoir à, être capable de, permettre de, etc.* ; anglais : *allow to, let, try to, start, etc.*), même si nous avons décrit les noms dépendant du verbe dont dépend syntaxiquement le verbe annoté comme participants, aucune fonction syntaxique directe n'a été indiquée. Nous avons tout de même noté le lien comme étant un lien indirect afin de ne pas perdre cette information importante du point de vue de la réalisation des participants. Par exemple, dans le contexte [5], le terme *load* dépend syntaxiquement du verbe *start*.

[5] *You'll see a boot menu before Windows starts LOADING.*

Dans ce cas, nous avons considéré que le participant *Windows* a un lien syntaxique indirect avec le terme *load*. En outre, nous indiquons qu'il réalise le rôle sémantique du Patient.

2.2 Rôles sémantiques

Une deuxième étape a consisté à proposer des équivalents anglais aux étiquettes de rôles sémantiques préalablement définies pour le français. Le tableau 2 liste quelques exemples de rôles en français avec leur équivalent anglais.

| Rôles sémantiques | | |
|-------------------|-------------|---|
| Français | Anglais | Exemples |
| Agent | Agent | You (Agent) INSTALL a card (Patient) in each PC (Location). |
| Destination | Destination | When you (Agent) LOAD the above html file (Patient) into a browser (Destination), then it displays the above tags as though it were formatted document. |
| Lieu | Location | You (Agent) INSTALL a card (Patient) in each PC (Location). |
| Manière | Manner | Windows (Agent) will now detect all plug-and-play devices and CONFIGURE them (Patient) automatically (Manner). |
| Moyen | Means | Each has a Web-based interface (Means) for ACCESSING files (Destination). |
| Patient | Patient | You (Agent) INSTALL a card (Patient) in each PC (Location). |

Tableau 2 : Exemples d'étiquettes de rôles sémantiques en français et en anglais.

Outre l'attribution d'équivalents anglais aux étiquettes préalablement définies pour le français, le travail d'adaptation à l'anglais a parfois exigé que nous proposons de nouveaux rôles sémantiques pour certains cas qui n'avaient pas encore

fait l'objet d'annotations en français⁴. C'est le cas des rôles *Criterion* et *Condition* illustrés en [6] et [7].

- [6] *The second is a program virus that (Assailant) INFECTS other programs (DESTINATION) when the infected program is run or executed (CONDITION).*
- [7] *In short, there is no « staging » of the backup process where storage and transmission resources (Patient) are ALLOCATED based on the priority of the data and the demand for access to it (Criterion).*

2.3 Fonctions et groupes syntaxiques

Une troisième étape a consisté à proposer des équivalents anglais aux étiquettes de groupes et fonctions syntaxiques, qui ne varient pas du français à l'anglais. Le tableau 3 ci-dessous liste les étiquettes de fonctions syntaxiques françaises et leurs équivalents anglais et le tableau 4, celles de groupes syntaxiques.

| Fonctions syntaxiques | |
|-----------------------|---------------|
| Français | Anglais |
| Sujet | Subject |
| Objet | Object |
| Complément | Complement |
| Modificateur | Modifier |
| Tête | Head |
| Lien indirect | Indirect link |

Tableau 3 : Liste des étiquettes de fonctions syntaxiques.

| Groupes syntaxiques | |
|------------------------------|---------------------------|
| Français | Anglais |
| SN (Syntagme Nominal) | NP (Noun Phrase) |
| SP (Syntagme Prépositionnel) | PP (Prepositional Phrase) |
| SAdv (Syntagme Adverbial) | AdvP (Adverbial Phrase) |
| Prop (Proposition) | Clause |
| SV (Syntagme Verbal) | VP (Verb Phrase) |
| Pro (Pronom) | Pro (Pronoun) |

Tableau 4 : Liste des étiquettes de groupes syntaxiques.

Les exemples d'annotation [8] et [9] montrent l'emploi des étiquettes de fonctions syntaxiques mentionnées dans les tableaux 3 et 4 :

⁴ Afin de conserver un parallélisme entre le français et l'anglais, les rôles ont immédiatement été ajoutés au schéma gérant l'annotation des contextes français.

[8] *If you have broadband and a small home network, you may even find that it helps you get online and ACCESS shared files more easily than you can with Windows.*

```
<fonction-syntaxique nom = "Indirect link">
  < groupe-syntaxique nom = "NP">
    you...
  </groupe-syntaxique>
</fonction-syntaxique>
get online and
< lexie-att> access </lexie-att>
<fonction-syntaxique nom = "Object">
  < groupe-syntaxique nom = "NP">
    shared files...
  </groupe-syntaxique>
</fonction-syntaxique>
<fonction-syntaxique nom = "Modifier">
  < groupe-syntaxique nom = "AdvP">
    more easily than you can with Windows...
  </groupe-syntaxique>
</fonction-syntaxique>
```

[9] *A username and password to be ENTERED before this system can be accessed.*

```
<fonction-syntaxique nom = "Head">
  < groupe-syntaxique nom = "NP">
    A username...
  </groupe-syntaxique>
</fonction-syntaxique>
and
<fonction-syntaxique nom = "Head">
  <groupe-syntaxique nom = "NP">
    password...
  </groupe-syntaxique>
</fonction-syntaxique>
to < lexie-att auxiliaire = "be"> be </lexie-att>
<lexie-att lemme = "enter"> entered </lexie-att>
...
```

3 Comparaison de quelques termes français-anglais et de leurs contextes annotés

Les annotations réalisées jusqu'à présent en anglais nous permettent déjà d'observer quelques différences dans le comportement des termes anglais et français souvent donnés comme équivalents exacts. Nous en donnons quelques-unes ci-dessous, basées sur l'observation des termes listés dans l'Annexe 1.

3.1 Structure actancielle

Certains actants associés aux rôles sémantiques entrant dans la composition de la structure actancielle des termes ne sont pas toujours réalisés de manière parallèle dans les contextes. Par exemple, les termes *click* et *cliquer* ont une structure

actancielle faisant appel aux rôles sémantiques Agent, Patient et Instrument. Dans les contextes anglais annotés jusqu'ici, les actants associés au rôle sémantique Instrument ne se réalisent jamais, tandis que dans les contextes français ils se réalisent dans trois cas (tableaux 5 et 6).

| CLIQUER 1 | | |
|------------|---------------------------|--|
| Actants | | |
| Patient | Complement (SP -sur) (12) | icône (3) bouton (3) ok celle-ci lien suivant onglet démarrer |
| Instrument | Complement (SP -avec) (3) | bouton (2) souris |
| Agent | Sujet (SN) (4) | vous (3) il |

Tableau 5 : Tableau récapitulatif du terme cliquer disponible en consultant le site du DiCoInfo (www.olst.ling.umontreal.ca/dicoinfo/).

- [10] Avec la souris, vous devez vous positionner sur l'option désirée et « CLIQUER » deux fois sur celle-ci.

| CLICK 1 | | |
|---------|--------------------------|---|
| Actants | | |
| Agent | Subject (NP) (4) | you (3) user |
| Patient | Complement (PP -on) (11) | button (2) symbol arrow option name link menu organize favorites item term |

Tableau 6 : Tableau récapitulatif du terme click disponible en consultant le site du DiCoInfo (www.olst.ling.umontreal.ca/dicoinfo/).

- [11] *Each time you click on this arrow, you will return to the next previous site that you visited.*

3.2 Classes sémantiques des actants

Les classes sémantiques des actants associés à certains rôles sémantiques peuvent varier du français à l'anglais pour des termes équivalents.

Par exemple, dans les contextes du verbe *connecter* les actants associés au rôle sémantique Agent peuvent être des animés ou des inanimés (tableau 7), tandis que dans les contextes de *log on* ils sont toujours des animés (tableau 8). Les contextes [12] et [13] illustrent ce cas de figure.

| CONNECTER 1 | | |
|-------------|---|--|
| Actants | | |
| Agent | Sujet (SN) (5) Lien indirect (SN) Sujet (Pro) (2) | vous (4) utilisateur (2) qui (l' utilisateur) qui (la machine) |
| Destination | Complement (SP -à) (12) | site (4) internet (3) système (2) serveur (2) réseau |

Tableau 7 : Comportement sémantique des participants associés au rôle sémantique Agent du terme connecter.

[12] Il sert à transporter le protocole TCP-IP sur une liaison temporaire, et permet à la machine qui se CONNECTE d'obtenir son adresse IP automatiquement.

| LOG ON 1 | | |
|-------------|---|--|
| Actants | | |
| Agent | Subject (NP) (9) Indirect link (NP) (7) | you (9) user (2) people (2) person i they (the bad guys) |
| Destination | Complement (PP -to) (11) | computer (3) internet (2) site (2) pc device application account |

Tableau 8 : Comportement sémantique des participants associés au rôle sémantique Agent du terme log on.

[13] *Telnet: This is a very insecure mechanism allowing remote users to LOG ON to your computer.*

La variation des classes sémantiques des actants associés à certains rôles sémantiques se vérifie également avec d'autres paires de termes comme: *visualiser₁* et *view₁*, *installer₂* et *install₂*, *lancer₁* et *launch_{1b}*, *générer₁* et *generate₁*, *écrire₁* et *write₁*.

3.3 Comportement syntaxique

Le comportement syntaxique des actants associés à certains rôles sémantiques varie du français à l'anglais. Les distinctions s'observent sur trois plans: celui des fonctions syntaxiques, des groupes syntaxiques et des options quant aux prépositions utilisées dans chaque langue.

3.3.1 Fonctions syntaxiques

Certains participants ayant le même rôle sémantique dans la même structure actancielle ont des fonctions syntaxiques différentes en français et en anglais.

Dans les contextes du terme *enregistrer*, les participants ayant le rôle sémantique Patient peuvent être des sujets ou des objets syntaxiques du verbe, tandis que dans les contextes anglais de *save* les participants désignant ce même rôle sémantique sont seulement des objets syntaxiques du verbe. Les tableaux 9 et 10 et les contextes [14] et [15] illustrent ceci.

| ENREGISTRER 1 | | |
|---------------|--|---|
| Actants | | |
| Patient | Sujet (SN) (12) Objet (SN) (6) | information (4) fichier (3) données (3) message photo bios document programme elles option paramètre |
| Destination | Complement (SP -sur) (13) Complement (SP -dans) (6) | disquette (4) fichier (4) mémoire (2) disque bande cartouche disque dur rom eprom support secteur périphérique |
| Agent | Sujet (SN) Lien indirect (SN) | ordinateur vous |

Tableau 9 : Comportement syntaxique des participants associés au rôle sémantique Patient du terme enregistrer.

- [14] Les fichiers sont ENREGISTRÉS sur des mémoires externes (disque, disquettes, bandes magnétiques, cartouches, etc.).

| SAVE 1 | | |
|-------------|--|---|
| Actants | | |
| Patient | Object (NP) (15) Object (Pro) (2) | file (6) document (4) that (a large, recently downloaded file) attachment update parameter data page it (the next data being sent to your computer) |
| Destination | Complement (PP -to) (3) Complement (AdvP) Complement (PP -in) Complement (PP -on) | location anywhere directory disk folder harddrive |
| Agent | Subject (NP) (3) Indirect link (NP) (2) Subject (Pro) | you (4) it (the cursor) computer |

Tableau 10 : Comportement syntaxique des participants associés au rôle sémantique Patient du terme save.

- [15] *If there is no hyperlink then you might check your cache folder for a large, recently downloaded file that you can SAVE TO ANOTHER DIRECTORY ON YOUR DISK.*

Des situations semblables s'observent dans d'autres paires de termes comme : *accéder₁* et *access₁*, *installer₁* et *install₁*, *lancer₁* et *launch_{1b}*, *download₁* et *télécharger₁*, *interpréter₁* et *interpréter₁*, *generate₁* et *générer₁*, *attach₁* et *joindre₁*.

3.3.2 Groupes syntaxiques

Certains participants ayant le même rôle sémantique et la même fonction syntaxique en français et en anglais peuvent, cependant, appartenir à des groupes syntaxiques différents.

Dans les contextes du verbe *joindre* et du verbe *attach*, les participants associés au rôle sémantique Destination sont toujours des compléments de leurs verbes respectifs. Toutefois, dans les contextes français ces participants peuvent se réaliser soit avec un syntagme prépositionnel soit avec un pronom, tandis que dans les contextes anglais ils peuvent se réaliser seulement avec un syntagme prépositionnel. Les tableaux 11 et 12 et les contextes [16] et [17] illustrent cette observation.

| JOINDRE 1 | | |
|-------------|--|--|
| Actants | | |
| Patient | Objet (SN) (13) Objet (Pro) Sujet (Pro) Tête (SN) (3) Sujet (SN) | fichier (11) note que {le fichier} qui {les virus} {les vers} .zip pièce élément document pièce jointe |
| Destination | Complément (SP -à) (8) Complément (Pro) | message (7) courrier y {un même courrier} |
| Agent | Sujet (SN) (5) | vous (5) |

Tableau 11 : Comportement syntaxique des participants associés au rôle sémantique Destination du terme joindre.

- [16] Mais en lisant ces pages, vous apprendrez aussi comment envoyer un même courrier à plusieurs personnes, y JOINDRE un document, créer un carnet d'adresses ou archiver les messages reçus dans des dossiers.

| ATTACH 1 | | |
|-------------|--|--|
| Actants | | |
| Patient | Subject (NP) (5) Object (NP) (8) | file (9) document (2) it identification |
| Destination | Complement (PP -to) (14) | message (11) posting email it |
| Agent | Subject (NP) (4) Indirect link (NP) (2) | you (5) user |

Tableau 12 : Comportement syntaxique des participants associés au rôle sémantique Destination du terme attach.

[17] *You can ATTACH more than one file to a single email message.*

Des situations semblables s'observent d'autres paires de termes comme: *installer₂* et *install₂*.

3.3.3 Options quant aux prépositions utilisées dans chaque langue

Des participants associés à certains rôles sémantiques sont des compléments des termes annotés et se réalisent dans les contextes français et anglais sous forme de syntagmes prépositionnels. Pourtant, la préposition utilisée varie d'une langue à l'autre (cette observation est valable tant pour les actants que pour les circonstants).

Par exemple, dans les contextes du terme *télécharger₁*, les actants associés au rôle sémantique Source sont des compléments du verbe et se réalisent sous forme de syntagmes prépositionnels régis par les prépositions *à partir de* et *sur*, tandis que dans les contextes anglais les compléments de l'équivalent *download₁* sont régis par la préposition *from*. De la même façon, dans les contextes français, les actants associés au rôle sémantique Destination sont des compléments du terme *télécharger₁* et se réalisent sous forme de syntagmes prépositionnels régis par les prépositions *sur* et *dans*, tandis que dans les contextes de *download₁*, ils sont régis seulement par la préposition *to*. Les tableaux 13 et 14 et les contextes [18] et [19] illustrent ceci.

| TÉLÉCHARGER 1 | | |
|---------------|--|--|
| Actants | | |
| Patient | Sujet (SN) (4) Objet (SN) (7) Objet (Pro) (2) Tête (SN) Lien indirect (SN) | fichier (6) version (2) célèbre que {des polices de caractères} logiciel police le (fichier) programme page |
| Destination | Complement (SP -sur) Complement (SP -dans) | disque dur répertoire |
| Agent | Sujet (SN) (4) Lien indirect (SN) | vous (3) usager on |
| Source | Complement (SP -à partir de) Complement (SP -sur) | ordinateur internet |

Tableau 13 : Choix de prépositions associés aux rôles sémantiques Source et Destination dans télécharger₁.

- [18] Vous pouvez TÉLÉCHARGER un petit programme sur Internet qui modifie votre modem 33,600 Kps en 55.
- [19] *This doesn't mean one should DOWNLOAD software from just any website.*

| DOWNLOAD 1 | | |
|-------------|--|---|
| Actants | | |
| Agent | Indirect link (NP) (2) Subject (NP) (8) | you (3) i (2) one we people person user |
| Patient | Object (NP) (17) Object (Pro) (2) Subject (NP) | file (6) program (3) version hwclear.exe java development environment software compiler driver linux that (the most common types of files) package that (popular technologies for creating programs) winzip |
| Source | Complement (PP -from) (6) | website (3) internet (2) site |
| Destination | Complement (PP -to) | directory |

Tableau 14 : Choix de prépositions associés aux rôles sémantiques Source et Destination dans download₁.

Des situations semblables se vérifient avec d'autres paires de termes comme: *installer₁* et *install₁*.

3.4 Autres

Certains participants associés à certains rôles sémantiques montrent une grande variété de réalisations dans une langue et moins dans l'autre.

Par exemple, les réalisations des participants associés au rôle Patient dans les contextes du terme *lancer₁*, ont une fréquence plus élevée que dans les contextes de son équivalent *launch_{1b}*. Ainsi, *application* et *programme* apparaissent cinq fois en français, tandis que ses équivalents anglais, *application* et *program*, apparaissent seulement une fois. Les tableaux 15 et 16 et les contextes [20] et [21] illustrent ce cas de figure.

| LANCER 1 | | |
|----------|---|---|
| Actants | | |
| Patient | Objet (SN) (15) Objet (Pro) Tête (SN) (2) | application (5) programme (5) logiciel (2) système (2) assistant le (le système d'exploitation) dos gestionnaire |
| Agent | Sujet (SN) (3) | vous (2) on |

Tableau 15 : Variation dans les réalisations associées au rôle sémantique Patient du terme lancer₁.

[20] Ouvrir l'attachement en LANÇANT l'application qui a servi à le créer.

| LAUNCH 1b | | |
|-----------|---|---|
| Actants | | |
| Agent | Indirect link (NP) (2) Subject (NP) (3) Complement (PP -by) | you (2) user digital's clientworks 2.9 windows it |
| Patient | Object (NP) (12) Subject (NP) (4) | macro operating system utility dialog program browser site application tool editor window they attachment windows 95 software copy |

Tableau 16 : Variation dans les réalisations associées au rôle sémantique Patient du terme launch_{1b}.

[21] *Boot up the computer, and it LAUNCHES Windows 95 and the TV software, and shifts into TV mode.*

Des situations semblables s'observent dans d'autres paires de termes comme : *insérer*₂ et *insert*₂.

Finalement, les annotations réalisées jusqu'à présent en anglais ont permis d'observer que certains termes en anglais n'ont pas d'équivalent lexical en français. C'est le cas de *launch*_{1a}. On observe une alternance inchoatif – causatif entre *launch*_{1a} (*the program launches*) et *launch*_{1b} (*the user launches a program*) qui fait défaut au français. Ainsi, *launch*_{1b} a comme équivalent le verbe *lancer*₁ (un utilisateur lance une application), mais le terme *launch*_{1a} demeure sans équivalent lexical véritable.

D'autres phénomènes de divergences interlinguistiques méritent d'être mentionnés. On observe en anglais une alternance entre les emplois transitifs et intransitifs de certains verbes. C'est le cas de *click*₁ qui s'emploie à la forme transitive (*click an icon*) et à la forme intransitive (*click on an icon*) sans qu'on soit en mesure de dégager une préférence nette pour l'un ou l'autre. L'équivalent français ne s'emploie qu'à la forme intransitive (cliquer sur une icône).

Conclusion

L'annotation des contextes anglais est encore à un stade initial. Rappelons que nous avons annoté les contextes associés à 25 termes anglais (associés à 25 équivalents français). Les contextes annotés peuvent être consultés en ligne. Nous comptons poursuivre le travail d'annotation afin de faire l'inventaire des différences de comportement entre les termes français et anglais.

Ce travail préalable révèle qu'il serait opportun de signaler les ressemblances et les divergences entre termes verbaux anglais et français en reliant formellement les contextes annotés et en les présentant de manière conviviale à l'utilisateur. Pour l'instant, les annotations dans chacune des langues sont affichées séparément. Une méthode comme celle proposée par Boas (2005) dans le cadre de l'extension de la méthodologie FrameNet à d'autres langues que l'anglais pourrait se révéler intéressante pour la terminologie également.

Enfin, nous pensons que ce travail pourra contribuer de manière générale à l'étude de l'équivalence en terminologie pour ce qui est des termes de nature prédicative.

Remerciements

Le travail présenté ici est financé par une subvention octroyée par le Conseil de recherches en sciences humaines (CRSH) du Canada. Nous tenons à remercier l'ensemble de l'équipe qui participe actuellement au projet : Fadila Hadouche, Marie-Ève Laneville, Guy Lapalme, Annaïch Le Serrec et Benoît Robichaud. Une partie du travail d'annotation des contextes français a été réalisé par Stéphanie Caron, Stéphanie Klebetsanis, Marie-Ève Laneville et Charlotte Tellier (2007-2008) que nous remercions également.

Bibliographie

- Binon (J.) *et al.*, 2000 : *Dictionnaire d'apprentissage du français des affaires*, Paris, Didier.
- Boas (H.C.), 2005 : « Semantic Frames as Interlingual Representations for Multilingual Lexical Databases », dans *International Journal of Lexicography*, vol. 18, n° 4, p. 445-478.
- DiCoInfo, 2009 : *Dictionnaire fondamental de l'informatique et de l'Internet*, Observatoire de linguistique Sens-Texte, www.olst.ling.umontreal.ca/dicoinfo/, consulté le 15 novembre 2009.
- Dolbey (A.), Ellsworth (M.) et Scheffczyk (J.), 2006 : « BioFrameNet: A Domain-Specific FrameNet Extension with Links to Biomedical Ontologies », dans Bodenreider (O.), éd., *Proceedings of KR-MED*, p. 87-94.
- Fillmore (C.), 1968 : « The case for case », dans Bach (E.) et Harns (R.), éd., *Universals in Linguistic Theory*, New York, Holt, Rinehard & Winston, p. 1-88.
- FrameNet, 2009 : *FrameNet*, dernière mise à jour le 20 août 2009, www.framenet.icsi.berkeley.edu/.
- Hadouche (F.) *et al.*, 2009 : « Intégration d'informations syntaxico-sémantiques dans les bases de données terminologiques : méthodologie d'annotation et perspectives d'automatisation », dans Amparo, (A.) et L'Homme (M.-Cl.), éd., *Proceedings of the First International Workshop in Terminology and Lexical Semantics (TLS'09)*, 19 June 2009, Université de Montréal, p. 22-31.
- L'Homme (M.-Cl.), 2008 : « Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés », dans *Traduire*, vol. 217, p. 78-103.
- L'Homme (M.-Cl.), Le Serrec (A.) et Laneville (M.E.), 2009 : *Guide d'annotation des contextes du DiCoInfo et du DiCoEnviro*, Université de Montréal, Observatoire de linguistique Sens-Texte (OLST).
- Lépinette (B.) et Cazevieille (F.), 2007 : « Un dictionnaire syntaxico-sémantique bilingue de la zoo-technie », dans *Cahiers de linguistique*, vol. 33, n° 1, p. 83-102.
- Ruppenhofer (J.) *et al.*, 2006 : *FrameNet II: Extended Theory and Practice*. www.framenet.icsi.berkeley.edu/, consulté le 15 novembre 2009.
- Schmidt, (T.) 2009 : « The Kicktionary – A multilingual lexical resource of football language », dans Boas (H.C.), éd., *Multilingual FrameNets in Computational Lexicography*, Berlin, New York, Walter de Gruyter, p. 101-134.
- Venturi, (G.) *et al.*, 2009 : « Towards a FrameNet Resource for the Legal Domain », dans Casellas, (N.) *et al.* éd., *Proceedings of the Third Workshop on Legal Ontologies and Artificial Intelligence Techniques, held in conjunction with the Workshop on Semantic Processing of Legal Texts, 8 June 2009, Barcelona, CEUR-WS*, vol. 465, www.ceur-ws.org/Vol-465/, consulté le 15 novembre 2009.
- VerbNet, 2009 : VerbNet, www.verbs.colorado.edu/~mpalmer/projects/verbnet, consulté le 15 novembre 2009.&

Annexe : Liste des termes annotés

| termes français | termes anglais |
|--------------------------|------------------------|
| accéder ₁ | access ₁ |
| accéder ₂ | access ₂ |
| charger ₁ | load _{ib} |
| cliquer ₁ | click ₁ |
| configurer ₁ | configure ₁ |
| connecter ₁ | log on ₁ |
| écrire ₁ | write ₁ |
| éditer ₁ | edit ₁ |
| entrer ₂ | enter ₂ |
| enregistrer ₁ | save ₁ |
| générer ₁ | generate ₁ |
| imprimer _{1a} | print _{1a} |
| infecter ₁ | infect ₁ |
| insérer ₂ | insert ₂ |
| installer ₁ | install ₁ |
| installer ₂ | install ₂ |
| interpréter ₁ | interpret ₁ |
| joindre ₁ | attach ₁ |
| lancer ₁ | launch _{ib} |
| naviguer ₁ | browse ₁ |
| surligner ₁ | highlight ₁ |
| taper ₁ | type ₁ |
| télécharger ₁ | download ₁ |
| traiter ₁ | process ₁ |
| visualiser ₁ | view ₁ |

François Maniez,
Centre de recherche en terminologie et traduction,
Université Lumière Lyon 2,
Lyon,
France,
francois.maniez@univ-lyon2.fr.

L'utilisation de l'adjectif relationnel dans les débats du Parlement européen : étude sur corpus portant sur son utilisation dans quatre langues romanes

Résumé

Après un bref rappel des propriétés des adjectifs relationnels, nous examinons leur emploi dans un corpus français de langue spécialisée avant de comparer l'usage de l'adjectif relationnel *décisionnel* avec ses équivalents en espagnol, italien et portugais dans le corpus des débats du Parlement européen. Les résultats obtenus montrent que les adjectifs relationnels sont plus fréquemment utilisés en italien et (dans une moindre mesure) en français que dans les deux autres langues étudiées.

Mots-clés : adjectif relationnel ; corpus multilingues ; langues romanes ; prémodification nominale ; traduction.

Introduction

Les adjectifs relationnels sont très fréquemment employés en terminologie dans les langues d'origine latine, et à un moindre degré dans les langues d'origine germanique. Un certain nombre d'adjectifs de la langue scientifique ou technique sont clairement des calques d'origine anglo-saxonne : ainsi, *behavioral*, dont l'ajout au lexique de l'anglais remonte à 1927, a vraisemblablement contribué à la naissance de l'adjectif français *comportemental* en 1949, même si d'autres calques morphologiques tardent à être adoptés par les lexicographes. Ainsi, selon le *Grand Dictionnaire terminologique*, les adjectifs *développemental* et *environnemental*, dont la naissance remonte respectivement à 1979 et à 1996 pour l'Office de la langue française du Québec, n'ont toujours pas franchi les portes du *Trésor de la langue française* et du *Dictionnaire de l'Académie française*, alors que les premières utilisations attestées de leurs équivalents anglais *developmental* et *environmental* remontent au début du xx^e siècle. La formation de nouveaux adjectifs français pour traduire des expressions équivalentes venant de l'anglais est cependant le plus souvent due à un phénomène syntaxique typique des langues germaniques, la prémodification nominale par un nom adjectival. La traduction par le complément de nom est parfois utilisée pour de telles structures, mais l'utilisation de l'adjectif

confère à l'expression utilisée une consonance plus technique par son aspect figé, tout en évitant le choix entre article défini ou indéfini pour introduire le complément de nom. Ainsi, *fracture reduction* peut être traduit par *réduction fracturaire* et *fracture site* par *foyer fracturaire*, alors que l'adjectif *fracturaire* n'a pas d'équivalent dans le lexique anglais (bien que *fracturaire* et ses divers composés (*anti-fracturaire*, *inter-fracturaire*, *micro-fracturaire*, *péri-fracturaire*, *post-fracturaire*, *pré-fracturaire*) totalisent une dizaine de milliers d'emplois sur la Toile, l'adjectif est absent du *Grand Robert*).

La structure du nom adjectival pose donc au traducteur et au terminographe dont le français est la langue cible le problème du choix entre deux structures possibles : celle du complément de nom et celle qui utilisera l'adjectif formé par dérivation à partir du nom français correspondant, si cet adjectif existe. Ce choix est souvent déterminé par l'usage : *cell repository* est traduit par *banque de cellules*, *cell wall* par *paroi cellulaire*, ces deux expressions n'admettant pas de variation. Quand un choix existe, il dépend également du locuteur et du contexte de communication. Un médecin écrira (et dira dans certains contextes) *cancer mammaire* ou *infarctus myocardique* là où le non-spécialiste parlera de cancer du sein et d'infarctus du myocarde.

Dans le cadre de cette étude, nous nous pencherons sur certaines propriétés des adjectifs dénominatifs (qui constituent la majorité des adjectifs dits relationnels), et nous examinerons dans un premier temps l'emploi de certains adjectifs relationnels dans un corpus français d'articles de recherche du domaine de la médecine totalisant vingt-trois millions de mots avant de comparer l'usage de l'adjectif relationnel *décisionnel* avec ses équivalents en espagnol, italien et portugais.

1 Caractéristiques morphologiques de l'adjectif relationnel

Il est souvent avancé que les adjectifs relationnels sont des dérivés nominaux (Mélis-Puchulu 1991 : 35) et leur non-prédicabilité est généralement attribuée à leur origine nominale. Selon Schuwer (2005), cet argument ne paraît pas valide, dans la mesure où tous les adjectifs relationnels ne sont pas dénominatifs (même si c'est la grande majorité), et où tous les adjectifs dénominatifs ne sont pas relationnels ou non prédicables. La définition des adjectifs relationnels selon le seul critère de non-prédicabilité (Bally 1944 : 96-97) a d'ailleurs été critiquée par de nombreux auteurs, notamment McNally & Boleda (2004). Pour Biber (1999), des adjectifs comme *additional*, *final* ou *similar* sont néanmoins considérés comme relationnels :

[...] relational classifiers (such as *additional*, *final*, *similar*) have little descriptive content, while many topical classifiers (such as *chemical*, *medical*, *political*) provide descriptive content while also limiting the reference of the head noun (Biber 1999 : 508).

Biber (1999 : 511) souligne que c'est principalement dans les écrits de type universitaire que de tels adjectifs sont utilisés, les plus fréquents étant selon lui *same*,

whole, different, full, general, major, final, main et *single*. On est donc loin d'une origine strictement dénominale dans cette définition. Nous pensons toutefois qu'il s'agit là d'une liste trop exhaustive et dont les propriétés sont difficiles à cerner. On observe notamment que le critère de dérivation nominale ne s'applique qu'à certains de ces adjectifs (*additional* est dérivé du nom *addition*, mais l'étymologie latine est nécessaire pour expliquer la dérivation de l'adjectif *final*, et les autres adjectifs cités ne sont pas dénominaux). Par ailleurs, les adjectifs que Biber nomme *topical classifiers* (*chemical, medical, political*) sont liés morphologiquement à des noms (*chemistry, medicine, politics*). L'appellation d'adjectif relationnel (*relational adjective*, ou plus rarement *relative adjective* en anglais chez certains auteurs) recouvre donc des réalités diverses en fonction des auteurs, d'où la tentation de recourir à des critères purement morphologiques pour définir cette catégorie.

2 Caractéristiques syntaxiques de l'adjectif relationnel

2.1 Non-prédicabilité

Comme le rappelle Schuwer (*op. cit.*) à propos de l'anglais (elle cite entre autres *alimentary, preventive* et *respiratory*), les mêmes adjectifs peuvent avoir des emplois relationnels et qualificatifs. Dans la mesure où de tels adjectifs sont déverbaux (soit par dérivation, soit en tant qu'emprunts, dans le cas d'*alimentary*), le critère de non-prédicabilité semble battu en brèche. Comme elle le fait justement remarquer,

[...] si ces qualificatifs ne sont pas prédicables lorsqu'ils sont associés à certains substantifs, c'est parce qu'ils contribuent à la création d'une notion complexe, perçue (et dite) comme type. La non-prédicabilité est la manifestation de cette création lexicale, qui explique que l'adjectif ne peut être désolidarisé du substantif.

Notons au passage qu'un test mentionné par Temple (1996) est le pendant de cette dualité d'emploi au niveau nominal quand on forme un substantif par suffixation en *-ité* à partir d'un adjectif dénominal. Ainsi, l'acceptabilité de l'expression « la cérébralité d'une peinture » par opposition à « *la cérébralité d'un lobe » est liée à la possibilité de prédication dans les emplois adjectivaux correspondants.

2.2 Non-gradation

La non-gradation semble être un critère un peu plus fiable que la non-prédicabilité, notamment l'incompatibilité avec l'adverbe « très ». Ainsi, « une production très laitière » semble difficilement acceptable, alors que la substitution d'autres adverbes (majoritairement, essentiellement) donne lieu à des séquences apparemment plausibles. Une piste intéressante pour l'anglais paraît être celle de la compatibilité avec *one*, comme le propose Ballier (à paraître). Partant de l'observation faite par Huddleston et Pullum (2002 : 1515) de l'impossibilité de remplacer par *one* certains noms recevant un complément introduit par *of*, il postule un autre type de contrainte permettant d'identifier un adjectif comme étant relationnel : la

nécessité, en cas d'anaphore par *one*, de la présence du co-référent dans le contexte amont immédiat à l'intérieur du même énoncé, les restrictions de sélection observées l'amenant à considérer les adjectifs relationnels comme une espèce de complément. Le test semble valide sur les données que nous avons analysées, et pourrait servir à confirmer ou infirmer le caractère relationnel de certains emplois adjectivaux, même si la relative rareté de telles structures en corpus rend probablement ce critère difficilement exploitable dans le cadre d'une extraction automatique.

3 Utilisations terminologiques de l'adjectif relationnel en médecine

On sait depuis L'Homme (2004) que les adjectifs dénominatifs constituent la majorité des adjectifs de la terminologie médicale. Comme le souligne Waltereit (2002), les relations entre les deux noms concernés par la prémodification en anglais (et entre l'adjectif et le nom en français dans les structures correspondantes) sont de natures multiples. Le prémodificateur peut signifier la localisation anatomique (*back pain* → douleur dorsale ou dorsalgie, *brain stem* → tronc cérébral), la fonction (*sweat glands* → glandes sudoripares), la cause (*heat rash* → erythème calorique) ou la forme (*sickle cell* → cellule falciforme). Ce peut également être un nom désignant une substance présente dans l'organisme. (*serum iron* → fer sérique, *plasma half-life* → demi-vie plasmatique).

3.1 Étude de la variation entre l'usage de l'adjectif relationnel et le complément de nom

Pour étudier les emplois des adjectifs relationnels français, nous avons eu recours à un corpus constitué d'articles de recherche dans le domaine médical tirés des revues disponibles sur la base de données *Science Direct*. Le corpus a été étiqueté en parties du discours par l'analyseur *Cordial*, et compte actuellement 23 millions de mots. Les diverses spécialités médicales sont à peu près également représentées, comme en témoigne la liste des revues utilisées: *Annales de Cardiologie et d'Angéiologie*, *Annales de Chirurgie*, *Annales de Chirurgie Plastique Esthétique*, *Annales Françaises d'Anesthésie et de Réanimation*, *Annales Médico-psychologiques*, *Annales de Réadaptation et de Médecine Physique*, *Annales d'Urologie*, *Médecine et Maladies Infectieuses*, *Néphrologie & Thérapeutique*, *Revue Française d'Allergologie et d'Immunologie Clinique*, *Revue de Médecine Interne*, *Revue du Rhumatisme*, *Transfusion Clinique et Biologique*. Nous ferons référence à ce corpus sous le nom de MED_FR dans les lignes qui suivent.

Nous avons comparé la fréquence des adjectifs relationnels modifiant le nom *risque* et des noms dont ils sont dérivés dans notre corpus, en examinant les emplois de ces noms comme complément pour tous les adjectifs relationnels de fréquence supérieure à 10. Les résultats apparaissent dans le tableau 1.

| Fréquence | Adjectif | Fréquence | Nom |
|-----------|----------------|-----------|--------------|
| 707 | hémorragique | 58 | hémorragie |
| 512 | infectieux | 271 | infection |
| 281 | fracturaire | 372 | fracture |
| 193 | allergique | 74 | allergie |
| 156 | suicidaire | 54 | suicide |
| 130 | thrombotique | 148 | thrombose |
| 47 | ischémique | 61 | ischémie |
| 39 | embolique | 57 | embolie |
| 23 | asthmatique | 166 | asthme |
| 16 | hémolytique | 19 | hémolyse |
| 13 | traumatique | 14 | traumatisme |
| 12 | anaphylactique | 38 | anaphylaxie |
| 11 | cicatriciel | 11 | cicatrice(s) |
| 11 | métabolique | 18 | métabolisme |

Tableau 1: Fréquence des adjectifs relationnels modifiant le nom risque et des noms dont ils sont dérivés dans le corpus MED_FR

L'observation de ces données montre une prédominance globalement marquée des adjectifs dans ce type de structure, seuls les adjectifs *asthmatique* et *anaphylactique* étant nettement moins utilisés que le complément de nom. À l'inverse, *hémorragique*, *infectieux*, *allergique* et *suicidaire* sont beaucoup plus utilisés que le complément de nom correspondant. Sur la Toile, que l'on considérera pour les besoins de la cause comme un corpus nécessairement moins spécialisé, les adjectifs relationnels sont systématiquement moins bien représentés que les compléments de nom. Ainsi, le moteur de recherche *Google* (interrogé le 10/09/2009) donne 25 900 pour *risque hémorragique* contre 127 000 résultats pour *risque d'hémorragie*. Dans les cas où l'emploi de l'adjectif est le fait des seuls spécialistes, le rapport peut dépasser 1/50 (5 140 pour *risque fracturaire* contre 326 000 pour *risque de fracture*). Cette différence entre les résultats observés dans notre corpus et sur la Toile montre clairement que l'adjectif relationnel est l'une des marques distinctives de l'article de recherche médicale.

4 Le cas de l'expression *processus décisionnel* dans le corpus *Europarl*

Lors de recherches antérieures (Maniez 2009), la terminaison en *-(i)onnel* était apparue productive pour les adjectifs relationnels français, et nous avons dans un premier temps étudié son emploi dans le corpus *Europarl*, corpus multilingue aligné des débats du Parlement européen. L'adjectif *décisionnel* étant le plus utilisé dans cette catégorie, nous avons examiné les équivalents de traduction de l'expression la plus fréquente du corpus qui contenait cet adjectif, soit la suite *processus décisionnel*, qui totalisait 83 % des emplois de l'adjectif. Les langues dont les équivalents de traduction ont été étudiés étaient l'anglais, l'allemand, l'espagnol, l'italien et le portugais.

4.1 Origine de l'expression « *decision-making process* » en anglais

S'il est relativement facile de dater les premières occurrences des lexies simples, cette datation s'avère plus difficile pour les expressions polylexicales non dictionnarisées. La première occurrence de l'expression anglaise « *decision-making process* » dans des écrits de type universitaire semble remonter à 1944 (elle est employée à plusieurs reprises dans un ouvrage de H.A. Simon publié à l'Université de Chicago, *Decision-making and administrative organization*). Dans le corpus en ligne de la revue *Time* (<http://corpus.byu.edu/time/>), on en relève une occurrence dans les années cinquante, puis 8 et 12 dans les deux décennies suivantes. L'expression est devenue d'usage extrêmement fréquent dans la deuxième moitié du xx^e siècle, et le moteur de recherche *Google* en recense actuellement plus de cent millions d'occurrences. Le TILF fait remonter à 1958 la première occurrence de l'adjectif *décisionnel* en français et cite le contexte « *Centres décisionnels. La « fonction décisionnelle » constitue la tâche majeure des détenteurs du pouvoir* » ; il est donc probable que *processus décisionnel* a rapidement été adopté dans les années soixante comme équivalent de traduction de l'expression anglaise, qui lui préexistait au moins d'une dizaine d'années. Notons au passage que l'adjectif relationnel *decisional* est d'un emploi extrêmement rare en anglais. Dans le corpus de la revue *Time*, l'adjectif ne compte que 2 occurrences, tout comme dans le *Grand Dictionnaire terminologique*. Pour ces deux dernières, il s'agit d'ajouts relativement récents : *decisional framework* (cadre décisionnel, 1987), et le synonyme *decisional architecture* donné pour *decision architecture*, (architecture décisionnelle, 1999).

4.2 Adjectifs français dérivés de *décision*

Outre l'adjectif *décisif*, le TILF donne *décisoire* dans son sens juridique (Qui détermine la décision d'un procès. *Serment décisoire*) et logique ou méthodologique (Qui apporte une solution définitive) et précise que l'on rencontre dans la documentation un adjectif considéré comme rare, *décisionnaire* (Qui tranche rapidement et d'autorité). L'adjectif *décisionnel*, (Qui ressortit à la décision) y est qualifié de néologisme, et il est attesté par le seul *Larousse de la Langue française*.

4.3 Étude détaillée du corpus

Sur 992 emplois de l'adjectif *décisionnel*, 821 concernent l'expression *processus décisionnel*. Un sondage portant sur dix occurrences a permis d'obtenir pour les langues étudiées les équivalents suivants (les abréviations suivantes seront désormais utilisées : FR pour le français, DE pour l'allemand, EN pour l'anglais, ES pour l'espagnol, IT pour l'italien, PT pour le portugais) : *Entscheidungsprozess* (DE), *decision-making process* (EN), *proceso de toma de decisiones* [6] et *proceso de decisión* [4] (ES), *processo decisionale* (IT), *processo de decisão* [8] et *processo de tomada de decisões* [2] (PT). Ce premier sondage fait apparaître une variation des équivalents de traduction utilisés en espagnol et en portugais, l'équivalent du verbe support *make* disparaissant dans les constructions en complément de nom de l'allemand *Entscheidungsprozess*, de l'espagnol *proceso de decisión* et du portugais *processo*

de decisão. Une exploitation manuelle de l'ensemble des occurrences montre une réalité un peu plus complexe, comme en atteste le tableau 2.

| Langue | Expressions utilisées |
|--------|---|
| FR | processus décisionnel [821], processus de décision [351], processus de prise de décision [145] |
| DE | Entscheidungsprozeß [823], Entscheidungsfindung [455], Beschlußfassung [383], Prozeß der Entscheidungsfindung [10] |
| EN | decision-making process [951], decision making process [20], decision-taking process [10] |
| ES | proceso de toma de decisiones [526], proceso de decisión [202], proceso decisorio [108], proceso de generación de decisiones [12] |
| IT | processo decisionale [1 524], processo di decisione [14] |
| PT | processo de decisão [735], processo decisório [278], processo de tomada de decisões [224] |

Tableau 2 : Fréquence des équivalents de traduction de l'expression *decision-making process* dans le corpus EuroParl

Les résultats font apparaître une forte variation des patrons syntaxiques des équivalents de traduction, à l'exception de l'italien, où l'adjectif relationnel *decisionale* est employé presque exclusivement (99 % des cas). En français, l'expression *processus décisionnel* est utilisée dans 62 % des cas, alors que les expressions espagnole (*proceso decisorio*) et portugaise (*processo decisório*) utilisant l'adjectif relationnel sont nettement moins fréquentes que celles utilisant un complément de nom (respectivement 13 % et 22 %). Le chiffre le plus parlant semble être le nombre extrêmement élevé d'occurrences de l'expression *processo decisionale* en italien, qui est beaucoup plus élevé que le nombre d'occurrences de l'expression anglaise employée pour la requête. Cette constatation nous a amené à examiner les contextes de ses utilisations pour lesquels *decision-making process* n'est pas présent dans l'équivalent anglais, ce qui donne les exemples suivants :

[1]

IT : *Ogni Stato membro dispone in linea di principio di un voto nel processo decisionale.*

EN : *Every member country has a vote in the decision procedure.*

FR : *Chaque État membre dispose en principe d'une voix dans le processus de décision.*

[2]

IT : *Nei processi decisionali le donne sono sottorappresentate ovunque.*

EN: *Women are universally under-represented where decisions are made.*

FR : Les femmes souffrent d'une sous-représentation universelle là où sont prises les décisions.

[3]

IT: *Siamo a conoscenza dei problemi comportati dal processo decisionale all'unanimità.*

EN: *After all, we know the problems that we see arising when decisions have to be taken unanimously.*

FR : Nous savons bien que les problèmes auxquels nous assistons surgissent lorsque les décisions sont prises à l'unanimité.

L'examen de ces exemples montre en [1] l'emploi de l'adjectif relationnel en italien comme équivalent de structures employant un complément de nom en anglais et en français, mais aussi, en [2] et [3], comme équivalent de structures dans lesquelles l'équivalent fait intervenir un verbe conjugué (*decisions have to be taken/ les décisions sont prises*). Les exemples [4] et [5] témoignent d'emplois de *processo decisionale* correspondant à l'expression anglaise *have/give a say*. Enfin, dans les exemples [6] et [7], on voit cette expression utilisée pour traduire la simple notion de politique, ou encore surajoutée à la mention du vote à la majorité qualifiée. On a donc l'impression d'une prédilection de l'italien pour ce type de structure.

[4]

IT: *Per un Parlamento che intende svolgere un ruolo significativo in seno alla Convenzione è dunque importante essere coinvolto nel processo decisionale sul bilancio.*

EN: *Then it is important, particularly for a Parliament which wants to play an important role in the Convention, to have a say in the budget.*

FR: Si le Parlement entend jouer un rôle important dans la Convention, il est essentiel qu'il ait son mot à dire dans la gestion du budget.

[5]

IT: *Ciò significa, a mio parere, maggiore partecipazione al processo decisionale, maggiore accesso alle informazioni, [...].*

EN: *By that I mean giving the individual a greater say, more access to information [...].*

FR: [...] autrement dit davantage de participation, un meilleur accès aux informations [...].

[6]

IT: *Questo è un altro punto debole del processo decisionale europeo.*

EN: *This is another weak area in our European policy.*

FR: Voilà un deuxième point sensible de notre politique européenne.

[7]

IT: *Perciò tale processo decisionale a maggioranza qualificata dovrà sempre essere abbinato al diritto di codecisione del Parlamento.*

EN: *That is why qualified majority voting must always go hand in hand with Parliament's right of codecision.*

FR: C'est pour cette raison que la majorité qualifiée doit toujours aller de pair avec le droit de codécision du Parlement.

Les observations que suscitent ces exemples nous amènent à formuler l'hypothèse selon laquelle l'italien utiliserait plus fréquemment l'adjectif relationnel que les autres langues romanes. Les résultats obtenus sur l'ensemble du corpus *Europarl* confirment cette hypothèse dans le cas des équivalents de l'adjectif *décisionnel*, employé 1 259 fois dans la partie française du corpus alors que l'italien *decisionale* l'est 3 282 fois (l'espagnol *decisorio* et le portugais *decisório* sont employés respectivement 348 et 396 fois, l'adjectif *decisional* étant d'usage encore plus rare dans ces deux langues, avec respectivement 7 et 9 occurrences). Cependant, ces résultats ne sont pas confirmés systématiquement lorsque d'autres familles d'adjectifs sont examinées. Ainsi, on obtient 4 938 occurrences pour *institutionnel*, 4 926 pour *institucional* (ES), 5 204 pour *istituzionale* (IT) et 4 868 pour *institucional* (PT).

Un autre critère peut être utilisé pour comparer les fréquences relatives d'emploi de l'adjectif relationnel dans les langues romanes: celui de la présence consécutive de deux d'entre eux à la suite d'un nom. Si l'on examine les 31 occurrences de l'expression « *processo decisionale comunitario* » dans la partie italienne du corpus, on observe seulement 5 occurrences de *processus décisionnel* (*au niveau*) *communautaire*, 1 occurrence de *proceso decisorio comunitario* (ES) et 1 occurrence de *processo decisional comunitário* (PT). La partie portugaise du corpus ne permet malheureusement pas de requêtes utilisant les parties du discours, mais la syntaxe < [pos = "NOM"] [pos = "ADJ.*"] [pos = "ADJ.*" & word = ".*onal.*"] >, qui isole les suites Nom Adjectif Adjectif de l'italien dont le deuxième adjectif est un dérivé en *-onal*, donne 6 701 résultats; la syntaxe équivalente donne 5 671 résultats pour l'espagnol et seulement 792 pour le français. Dans le cas des adjectifs se terminant en *-taire*, les chiffres sont de 2 509 pour l'italien (syntaxe: < [pos = "NOM"] [pos = "ADJ.*"] [pos = "ADJ.*" & word = ".*tari.*"] >), 1991 pour l'espagnol, et 2 439 pour le français. Dans le cas des adjectifs se terminant en *-atif*, les chiffres sont de 2 352 pour l'italien, 2 035 pour l'espagnol, et 2 236 pour le français (syntaxe: < [pos = "NOM"] [pos = "ADJ.*"] [pos = "ADJ.*" & word = ".*ati [fv].*"] > pour ce dernier cas). On obtient donc des chiffres régulièrement plus élevés du côté de l'italien et dans une moindre mesure du français, même si la fréquence de l'emploi semble varier entre le français et l'espagnol en fonction des suffixes adjectivaux.

Dans la mesure où l'expression à l'origine de cette étude (*decision-making process*) correspond à un patron syntaxique récurrent en anglais (N-Ving N), nous avons examiné une deuxième hypothèse, celle selon laquelle ce type de structure favoriserait l'émergence des adjectifs relationnels dans les équivalents de traduction des langues romanes. L'étude d'expressions contenant des adjectifs tels que *life-saving* ou *life-threatening* ne donne cependant pas de résultats probants. Le phénomène le plus souvent observé est en fait le suivant : l'absence de dérivation adjectivale favorise majoritairement l'emploi de propositions relatives, comme dans le cas des traductions utilisées dans les langues romanes du corpus pour l'expression *competition-distorting domestic support* dans l'exemple [8] :

[8]

EN: [...] *substantial reductions in competition-distorting domestic support.*

ES: [...] *reducciones esenciales en el campo de las medidas de apoyo que distorsionan la competencia.*

FR: [...] *diminutions considérables des mesures de soutien qui faussent le jeu de la libre concurrence.*

IT: [...] *sostanziali riduzioni nel settore delle misure di sostegno che determinano distorsioni della concorrenza.*

PT: [...] *reduções substanciais no domínio das medidas de apoio que comportam distorções da concorrência.*

Il est donc probable que le phénomène soit limité aux équivalents de traduction des structures faisant intervenir un verbe support comme *decision-making*, ce qui expliquerait le passage à deux unités de sens à la traduction d'expressions comme *decision-making process*. On observe ce mécanisme pour le français l'italien et le portugais dans l'exemple [9] avec l'expression *constitution-making process* :

[9]

EN: [...] *the same month as the decisive moment in the constitution-making process.*

ES: [...] *el mismo mes decisivo para el proceso de redacción de la Constitución.*

FR: [...] *le même mois que pour le moment décisif du processus constitutionnel.*

IT: [...] *mese che è stato decisivo anche per l'attuale processo costituente.*

PT: [...] *o mesmo mês do momento decisivo neste processo constituinte.*

Conclusion

L'italien et le français semblent utiliser l'adjectif relationnel plus fréquemment que l'espagnol et le portugais, dans des proportions restant à déterminer. Il reste également à savoir si cette différence de fréquence concerne davantage certains types de suffixation, voire certains adjectifs isolés. Quant aux adjectifs anglais de patron syntaxique N-V + *-ing*, ils sont susceptibles d'être traduits par des adjectifs

relationnels dans les langues romanes quand ils font intervenir un verbe support, les autres cas de figure donnant le plus souvent lieu à une décomposition des adjectifs anglais sous la forme d'une proposition relative.

Bibliographie

- Ballier (N.), à paraître : *The semantics of one-substitution test*, Manuscrit.
- Béjoint (H.) et Maniez (F.), 2005 : *De la mesure dans les termes*, Lyon, Presses Universitaires de Lyon.
- Coates (J.), 1971 : « Denominal adjectives: A study in Syntactic relationships between Modifier and Head » dans *Lingua* 27, p.160-169.
- Daille (B.), 2001 : « Qualitative terminology extraction: Identifying relational adjectives », dans Bourrigault (D.), Jacquemin (C.), et L'Homme (M.-Cl.), éd., *Recent advances in computational terminology*, Amsterdam, John Benjamins Publishing Company, p.149-166.
- Fradin (B.), 2008 : « Les adjectifs relationnels et la morphologie », dans *La Raison Morphologique: hommage à la mémoire de Danielle Corbin*, Amsterdam, John Benjamins Publishing Company, p.69-91.
- L'Homme (M.-Cl.), 2004 : « Adjectifs dérivés sémantiques (ads) dans la structuration des terminologies », dans *Actes de Terminologie, ontologie et représentation des connaissances*, Université Jean-Moulin Lyon-3, p.22-23 janvier 2004.
- Maniez (F.), 2005 : « Identification automatique des adjectifs relationnels: une étude sur corpus », dans Béjoint (H.) et Maniez (F.), éd., *De la mesure dans les termes*, Lyon, Presses Universitaires de Lyon, p.134-152.
- Maniez (F.), 2009 : « La mise en équivalence des adjectifs relationnels du domaine médical: étude du suffixe *ionnel* », dans Alcina (A.) et L'Homme (M.-Cl.), éd., *Actes du Workshop international sur la Terminologie et la sémantique lexicale (TLS'09)*, Université de Montréal, juin 2009, p.32-41, olst. ling. umontreal. ca/pdf/ProceedingsTLS09.pdf.
- McNally (L.) et Boleda (G.), 2004 : « Relational adjectives as properties of kinds », dans Bonami (O.) et Cabredo Hofherr (P.), éd., *Empirical Issues in Formal Syntax and Semantics* 5, p.179-196, www.cssp.cnrs.fr/eiss5
- Mélis-Puchulu (A.), 1991 : « Les adjectifs dénominaux: des adjectifs de relation », dans *Lexique*, 10, p.33-60.
- Schuwert (M.), 2005 : « Systèmes des adjectifs non prédicatifs en français et en anglais: constantes et variations », dans *Cahiers de lexicologie*, vol. 86, p.85-103.
- Temple (M.), 1996 : *Pour une sémantique des mots construits*, Villeneuve d'Ascq, Presses Universitaires du Septentrion.
- Waltereit (R.) 2002 : « Le rapport dépendanciel entre adjectif et nom: données syntaxiques et structures conceptuelles », dans *Syntaxe et Sémantique*, 4, Caen, p.179-194.

Corpus

Time Magazine Corpus of American English: corpus.byu.edu/time/

Corpus de français médical du Centre de Recherche en Terminologie et Traduction: perso.univ-lyon2.fr/~maniezf/Corpus/Corpus_medical_FR_CRTT.htm

Sources lexicographiques

Dictionnaire de l'Académie française: www.academie-francaise.fr/dictionnaire/

Grand Dictionnaire terminologique: www.granddictionnaire.com/

Grand Robert de la langue française, 2007, Paris, Le Robert.

Trésor de la langue française informatisé: atilf.atilf.fr/

Séance de clôture

Allocution de clôture

Marc Cheymol,
Directeur,
Langue française, diversité culturelle et linguistique,
Agence universitaire de la Francophonie

Madame la Rectrice honoraire de l'AUF,

Madame la Professeure Teresa Lino,

Monsieur le Coordonnateur du Réseau LTT,

Mesdames et Messieurs,

Chers Collègues, Étudiants, Enseignants et Chercheurs,

Je le dirai sans ambages : les *Huitièmes Journées scientifiques LTT* qui s'achèvent aujourd'hui m'apparaissent comme un grand succès.

Pour ce succès je voudrais remercier une nouvelle fois l'Université nouvelle de Lisbonne, ses autorités qui ont apporté leur soutien et qui nous ont offert un si convivial dîner hier soir, ses grandes dames et ses petites mains, tous les enseignants-chercheurs qui se sont transformés en organisateurs pour nous accompagner et s'occuper de nous au cours de ces trois journées, dans ce lieu sublime et confortable dont il faut aussi saluer le Président pour son accueil impeccable.

En disant que les Journées scientifiques sont un grand succès, j'ai parfaitement conscience de ne rien dire d'original, je crois au contraire exprimer assez platement le sentiment le plus largement répandu. Mais je souhaiterais ici donner quelques arguments qui l'expliquent et le justifient. Ces arguments résultent des journées elles-mêmes, du Réseau LTT et du contenu scientifique de cette rencontre.

1. En ouvrant cette manifestation, j'ai mentionné la diversité géographique, linguistique, disciplinaire des participants. Celle qui me frappe aujourd'hui au moment de les conclure, c'est la diversité générationnelle et c'est sur elle que je souhaiterais insister en premier lieu.

J'ai assisté à bien des journées scientifiques, y compris du Réseau LTT, et ce qui me surprend, dans celles-ci, c'est le renouvellement des personnes. C'est la force d'entraînement que semble exercer LTT dans la mise en œuvre de ses journées. Je parle ici devant des spécialistes de la métaphore, que vous avez analysée par exemple dans le langage médical, et cependant je me risque à employer ici le mot *entraînement*, non pas comme vocabulaire du sport (même si les journées scientifiques ont indiscutablement quelque chose de sportif!), mais comme un terme de mécanique, au sens où LTT apparaît comme une sorte de machine savamment agencée avec des courroies et des poulies, des leviers de transmission et des engrenages disposés de telle sorte qu'ils assurent eux-mêmes le remplacement progressif de leurs différentes pièces.

Les JS sont un véritable outil de ce renouvellement. En donnant aux jeunes chercheurs qui les fréquentent la possibilité de présenter leurs premières recherches, en encadrant leurs exposés, en leur prodiguant avertissements et conseils, vous contribuez indiscutablement à l'objectif – que j'ai cité en commençant comme une priorité de l'AUF – de faire émerger une nouvelle génération d'experts et de chercheurs dans votre domaine.

2. Le deuxième élément de ce succès, c'est LTT. Un mot qui se suffit à lui-même, bannière trilobée à laquelle vous avez réussi à donner une existence internationale. LTT a d'abord été un « réseau » de l'AUF. J'ai évoqué la réforme à laquelle réfléchit l'AUF, comme l'occasion, pour LTT de passer à autre chose. Ce que j'ai entendu au cours de ces journées me donne bon espoir de voir LTT sous sa forme actuelle évoluer vers une autre forme d'organisation qui conserve les atouts dont nous venons de constater les effets, comme un groupement de laboratoires, par exemple, qui serait complémentaire du répertoire des chercheurs qui dans l'AUF regroupera les individus en collectifs autour de grands pôles disciplinaires comme un pôle Langues.

Dans le cadre de cette réforme, les actions de recherche que vous connaissez, dont nous avons eu la présentation des quatre projets en cours, seraient remplacées par des projets de plus grande envergure, d'une durée de quatre ans, qui seraient capables de combiner les différents outils de l'AUF dans des dynamiques plus ambitieuses intégrant la recherche et la formation à la recherche, des formations à distance, des mobilités, des publications. Dans cette perspective les journées comme celles-ci devraient aussi devenir autre chose, un élément de l'un de ces projets combinant, comme c'était déjà le cas ici, les journées de formation et les journées scientifiques, mais en les axant encore plus directement sur une thématique et sur des besoins liés aux priorités dans ce domaine.

Ce « rite de passage » ne devrait pas être trop difficile dans la mesure où, comme l'a exprimé votre coordonnateur dans la séance inaugurale, vous avez souhaité ces changements et vous les avez d'une certaine manière devancés.

3. Sur le plan du contenu de cette rencontre, deux grands thèmes ont appelé mon attention et me semblent capables de fédérer des projets de ce type: l'intercompréhension et les bases de données.

Il s'agit de deux termes – ce sera ma première remarque – qui recouvrent des réalités très diverses, et qui sont investis de sens différents selon la personne qui les emploie :

- à partir des exemples qui ont été donnés, nous avons constaté que l'intercompréhension pouvait jouer entre des langues qui présentent des similitudes, ce qui permet d'en apprendre – au moins pour une compréhension passive – plusieurs à la fois : entre les langues romanes, ou entre les langues slaves. Mais vous avez aussi relevé une possibilité d'intercompréhension qui serait viable sur le plan pédagogique, entre le français et une langue très différente comme l'amazigh. L'élaboration des méthodes et des manuels qui pourraient utiliser ces possibilités, et qui permettraient leur introduction dans les curriculums, même dans les écoles, ce qui soutiendrait alors un projet politique de développement de l'intercompréhension.
- le second thème a fait l'objet de nombreuses communications. Les bases de données peuvent recouvrir plusieurs réalités à plusieurs niveaux : nous avons vu des exemples de bases de données conçues dans le cadre d'un projet de thèse individuelle, sur *Access*, comme des projets plus ambitieux, qui utilisent des logiciels professionnels placés sur un serveur. Vous avez ressenti la nécessité d'une sorte de guide ou de portail qui permettrait aux chercheurs de s'orienter dans toutes ces bases de données expérimentales ou professionnelles, et qui proposerait aussi des outils pour les réaliser.

Ainsi, ces deux thèmes fédèrent des chercheurs qui pourraient se retrouver transversalement dans ce que nous avons proposé comme des collectifs : l'intercompréhension est l'un des thèmes prioritaires du *Collectif Didactiques* ; les bases de données feraient l'objet d'un nouveau *collectif sur les bases de données*, que nous bâtirons avec tous ceux qui s'y intéressent.

Dans un cas comme dans l'autre, nous avons constaté aussi l'attention que vous portez à la variation des langues. Tout d'abord, la variation des usages du français selon la géographie de la francophonie. Ainsi, vous avez donné des exemples de bases de données qui permettraient de passer de simples dictionnaires bilingues à des « dictionnaires bi-langues », donnant les équivalents d'un mot français dans tous ses sens différents selon le pays où il est parlé, mais aussi de la variation dans les langues cibles ou sources, autres que le français : les portugais européen et brésilien, le portugais du Cap-Vert, et celui de l'Angola, etc.

Vous avez aussi montré comment les bases de données linguistiques se constituent comme une sorte de gigantesque Encyclopédie, un grand réservoir où seront répertoriées et organisées toutes les connaissances sur les langues, par exemple les termes qui existent dans les langues mais pas encore dans les dictionnaires, ou

des catégories de termes qui existent dans d'autres langues d'une même famille. La constitution de bases de données va donc au-delà de la description des langues et devient une source de toutes sortes de dictionnaires, bilingues, ou analogiques, ou étymologiques, voire de manuels de conjugaison, par exemple.

Pour conclure, je voudrais insister sur le fait que vos recherches ne se présentent donc pas comme des recherches isolées, autarciques, faites pour le plaisir de la recherche. Vous avez présenté des recherches qui décrivent le lien entre la langue des spécialistes et les usages sociaux de la langue ou des recherches qui résultent de commandes des entreprises, dans deux domaines qui contribuent au développement: la médecine, qui assure la santé des personnes, et la finance qui assure la santé de leurs biens. Comment un pays peut-il se développer si les malades ne comprennent pas les médecins, si les entreprises n'ont pas accès aux outils de la langue?

C'est le rôle de la langue dans la qualité de la communication et dans la qualité de la pensée qui est en jeu.

Mais « tout n'est pas pour le mieux dans le meilleur des mondes », même si l'expression chère à Pangloss a été employée par l'un d'entre vous, et à juste titre puisqu'une scène de *Candide*, comme vous le savez, se passe justement à Lisbonne. Nous avons entendu se manifester des inquiétudes sur la généralisation d'une hyper-langue mondialisée qui est imposée à la recherche, même à la recherche sur le français, et dont le véritable danger, me semble-t-il, est comme vous l'avez remarqué, qu'elle est de mauvaise qualité. En effet, je ne crois pas qu'il faille plaider pour une défense du français pour le français, mais pour que la recherche se fasse et s'exprime avec précision et avec justesse, non dans une langue approximative ou mal contrôlée, car c'est la qualité de la recherche qui est en jeu.

Tel est le défi qui nous a été lancé. C'est le défi de la diversité.

Je vous remercie de l'avoir exprimé dans des termes précis et d'avoir ainsi montré que la recherche en langues peut contribuer à faire émerger une nouvelle génération de chercheurs *acteurs du développement*.

Je vous remercie.

Table des matières

| | |
|---|-----|
| Index des auteurs | 5 |
| Séance d'ouverture | 9 |
| Communications | 21 |
| Aleksandra Dziadkiewicz, <i>Vers un dictionnaire pragmatique français-polonais, polonais-français : quels critères pour le choix des entrées ?</i> | 23 |
| Mathieu Mangeot, Hong-Thai Nguyen, <i>Projet MotÀMot : élaboration d'un système lexical multilingue par le biais de dictionnaires bilingues</i> | 35 |
| Valeria Zotti, <i>Pour une réinterprétation du dictionnaire bilingue face à la pluralité linguistique de l'espace francophone : l'exemple du français québécois</i> | 49 |
| Samia Bouchaddakh, <i>Le Dico-FRAR : base de données lexicographiques bilingue français-arabe</i> | 63 |
| Miloud Taïfi, Patrice Pognan, <i>Le Dictionnaire raisonné berbère-français. Parlers du Maroc et ses traitements informatiques</i> | 83 |
| Jorge Baptista, Dolors Català, <i>Adverbes focalisateurs et analyse syntaxique automatique de groupes nominaux</i> | 97 |
| Carla Serhan, <i>La traduction de la forme pronominale réfléchie en arabe et en anglais : révélation du rapport au corps</i> | 111 |
| Zrinka Šimunić, <i>Analyse comparative du discours comme méthode d'évaluation de la qualité des productions textuelles lors d'une épreuve de traduction écrite</i> | 123 |
| Kaja Goštkowska, <i>Les terminologies française et polonaise du génie biomédical face au défi de la diversité</i> | 139 |
| Dardo de Vecchi, <i>Les mots des entreprises : des termes mal connus</i> | 149 |
| Jeanne Dancette, <i>Un dictionnaire encyclopédique plurilingue sur thésaurus</i> | 161 |

- Pierre Lerat,**
*Communautés de travail et diversité terminologique :
le préemballage, ses conceptualisations et ses enjeux* 177
- Taoufik Massoussi, Gérard Petit,**
Métalangue, une base de données multilingue 185
- Salah Mejri,**
La traduction d'un dictionnaire des sciences du langage : étude de cas 197
- Manuel Célio Conceição,**
Lexicologie et terminologie face aux défis de la communication multilingue 211
- Alexandra Feldekircher Müller, Maria da Graça Krieger,**
*Reconnaissance de la terminologie du droit de la protection sociale dans l'Arrêt
juridique sur la protection sociale* 219
- Françoise Bacquelaïne,**
*Technologies de pointe, terminologies et diversité linguistique :
la terminologie Bluetooth en français et en portugais* 235
- Taoufik Massoussi, Soumaya Mejri,**
Traitement automatique des discours financiers : application à la bourse 245
- Mathieu Guidère,**
Vers un dispositif interactif d'apprentissage de la traduction à distance 259
- Abdallah El Mountassir,**
*Vers un modèle d'intercompréhension progressive entre l'amazigh et le français.
L'expérience de l'enseignement de l'amazigh à l'université marocaine* 277
- Anna Anastassiadis-Syméonidis, Maria Mitsiaki, Valentina Chouma, Eleni Vletsi,**
*Analyse d'erreurs de langue des apprenants du grec
langue seconde dans un corpus écrit* 287
- Liliana Alic,**
Réussir la traduction 299
- Lamria Chetouani,**
*Le français transversal vecteur des enseignements/apprentissages
des disciplines scientifiques* 311
- Christine Jacquet-Pfau, Jean-François Sablayrolles, John Humbley,**
Emprunts, créations « sous influence » et équivalents 325
- Sandra de Caldas,**
*Néologie terminologique et emprunt : résultat d'une interculturalité qui défie le temps.
Analyse et traitement des emprunts lexicaux en français et en portugais
dans un corpus officiel bilingue spécialisé dans le domaine de l'enseignement* 341

| | |
|--|-----|
| Alicja Kacprzak, <i>Diversité technolœctale en diachronie: le cas de quelques termes œdicaux franœais</i> | 355 |
| Ieda Maria Alves, <i>Nœologie et sociœtœ mondialisœe: implications morphologiques</i> | 367 |
| Mirabela Navlea, Amalia Todiraœcu, <i>Ressources linguistiques pour un outil de traduction automatique statistique factorisœe franœais-roumain</i> | 375 |
| Serge Verlinde, <i>La Base lexicale du franœais: un outil lexicographique multi-usage</i> | 389 |
| Patrice Pognan, <i>Bases de donnœes pour l'œtude d'un groupe de langues: l'exemple des langues slaves de l'Ouest</i> | 399 |
| Mercœ Vazquez Garcia, Antoni Oliver Gonzœlez, <i>Utilisation de stratœgies statistiques pour la rœcupœration automatique de termes</i> | 419 |
| Isabelle Oliveira, Sabine Ploux, <i>Vers une mœthode de dœtœction et de traitement automatique de la mœtaphore</i> | 429 |
| Patrick Drouin, Frœdœric Doll, <i>Potentiel terminologique, quel sens prendre ?</i> | 441 |
| Mosbah Saïd, <i>Traduire les thœories linguistiques: l'exemple de la psychosystœmatique du langage</i> | 455 |
| Lucie Raharinirina Rabaovololona, <i>Traduction des expressions œ emprunt vers leur langue source, un rite de passage de la diversitœ linguistique et culturelle</i> | 469 |
| Andrœ Valli, Alberto Bramati, <i>Traduire les mœtaphores ontologiques dans l'espace roman. Le cas du franœais vers l'italien</i> | 479 |
| Van Dai Vu, <i>Mœtaphore conceptuelle en traduction juridique et œquivalence</i> | 493 |
| Pierre J.L. Arnaud, <i>Dœtœcter, classer et traduire les mœtonymies (anglais et franœais)</i> | 503 |
| Pedro Mogorrœn Huerta, <i>Compœtence phrasœologique et traitement des expressions figœes dans les dictionnaires</i> | 517 |

- Abdelaziz Berkai,**
*Quelques problèmes lexicographiques que posent l'établissement d'équivalences
 et leur organisation dans un dictionnaire kabyle-français* 537
- Heidi Verplaetse, Cornelia Wermuth, Philippe Anckaert, June Eyckmans, Winibert Segers,**
*La qualité de la traduction médicale
 et les exercices terminologiques préparatoires* 549
- Madalena Contente, Sebastião Silva Filho,**
Ontoterminologie de l'éducation médicale: conception et organisation 557
- Marie Calberg-Challot, Christophe Roche, Luc Damas,**
*Définition du terme vs définition du concept:
 pour une diversité langagière et une conceptualisation commune* 567
- Natalie Kübler, Geneviève Bordet, Mojca Pecman,**
*La linguistique de corpus entretient-elle d'étroites relations
 avec la traduction pragmatique?* 579
- Rosa Maria Fréjaville,**
*Le corpus en analyse terminologique de domaines composites
 et interdisciplinaires* 593
- Lorraine Goeuriot, Yukie Nakao, Béatrice Daille,**
Analyse de la modalité dans un corpus spécialisé multilingue 605
- Emmanuel Cartier,**
Repérage automatique des contextes définitoires dans les textes 619
- Hélène Ledouble, Renate Trurnit Verbic,**
*Le « Vert »: couleur, signal, symbole.
 Étude lexicographique à partir d'un corpus quadrilingue* 633
- Janine Pimentel, Marie-Claude L'Homme,**
*Annotation syntaxico-sémantique de contextes spécialisés:
 application à la terminographie bilingue* 651
- François Maniez,**
*L'utilisation de l'adjectif relationnel dans les débats du Parlement européen:
 étude sur corpus portant sur son utilisation dans quatre langues romanes* 671
- Séance de clôture** 683

Achévé d'imprimer

Au-delà d'une valorisation scientifique de leurs recherches, les lexicologues, terminologues et traductologues sont animés par une volonté farouche de décrire les langues, d'en connaître mieux les fonctionnements pour les rendre aptes à exprimer la modernité. L'homme ne peut se développer, comprendre le monde, acquérir des connaissances, adhérer à des valeurs universelles et faire partager les siennes propres que si sa langue maternelle est instrumentée pour ce faire.

La question de la transmission des connaissances et de l'intercompréhension entre toutes les langues en présence dans les aires linguistiques des chercheurs francophones, hispanophones, lusophones ou anglophones est loin d'être résolue en dépit des innovations méthodologiques et technologiques. En dépassant le cadre de la seule francophonie plurilingue, les Journées scientifiques LTT de Lisbonne avaient pour objectif d'offrir un espace d'échanges pluridisciplinaires entre chercheurs confirmés pour revisiter, dans leurs champs d'investigations respectifs, les théories et paradigmes autour du multilinguisme. Ces actes publient les meilleures communications retenues par le comité scientifique.

Depuis vingt ans, le Réseau Lexicologie, terminologie, traduction (LTT) inscrit ses travaux au cœur de la problématique de la diversité linguistique et du plurilinguisme. Il rassemble des chercheurs issus d'une grande diversité de pays francophones, voire de territoires où le français est une langue seconde, sans statut officiel ni lien génétique ou historique avec les langues locales.

La communauté des chercheurs LTT étudie des questions de langues qui constituent autant de facettes de la diversité linguistique. Celle-ci peut, en effet, concerner aussi bien des problématiques de variations codiques internes que des problématiques de contacts interlinguistiques. Ces recherches, parfois fortement marquées par l'ingénierie, vont de pair avec un souci de préservation, sinon de développement de la diversité. En témoigne une longue tradition d'actions de recherche orientées vers les langues du Sud et nombre d'engagements personnels en faveur de ces langues.

Prix public: 52 euros

Prix spécial membres AUF dans les pays en voie de développement : 30 euros HT

ISBN : 9782813000521



eo
ac
éditions des
archives
contemporaines

AGENCE
UNIVERSITAIRE
DE LA FRANCOPHONIE