

OLST-FQRSC
Rapport d'activités
mai 2004 – avril 2006

Vers une approche unifiée
de la lexicographie de langue générale
et de la terminologie multilingue

Table des matières

1.	Introduction.....	3
2.	Projets et état d'avancement des travaux.....	3
2.1	Dictionnaires et bases de données lexicales	4
2.2	Analyses de données lexicales.....	5
2.3	Modélisation de données lexicales	6
2.4	Corpus.....	7
2.5	Traitement automatique de la langue.....	7
3.	Personnel.....	9
3.1	Professionnels de recherche.....	9
3.2	Étudiants	9
3.3	Stagiaires.....	10
4.	Autres activités	11
4.1	Séminaire, colloques.....	11
4.2	Professeurs, chercheurs invités	11
4.3	Collaborations.....	12
5.	Publications.....	12
a.	Livres ou chapitres de livres	12
b.	Articles dans revues avec comité de lecture	13
c.	Articles dans actes de congrès	14
d.	Communications non publiées.....	16
e.	Autres.....	17

1. Introduction

La lexicographie et la terminologie visent toutes deux la modélisation d'un même type d'objets : les lexies (ou unités lexicales). Pour la lexicographie, il s'agit avant tout des lexies de la langue générale, décrites dans des dictionnaires (papier ou en ligne) et dans des bases de données lexicales. Pour la terminologie, il s'agit des lexies de langues de spécialité, souvent considérées selon une perspective plus « conceptuelle » que les lexies de langue générale, notamment en terminologie multilingue. Bien qu'étroitement liées par leur objet d'étude, les deux disciplines en question sont la plupart du temps pratiquées de façon autonome, sans que les linguistes, traducteurs et terminologues impliqués cherchent véritablement à unifier leurs approches théoriques et méthodologiques.

L'Observatoire de linguistique Sens-Texte (OLST) présente l'originalité de regrouper ces trois catégories de chercheurs, qu'unit un même intérêt pour la modélisation formelle des lexiques. Sans se situer tous dans le strict cadre notionnel de théorie linguistique Sens-Texte (qui a donné son nom au groupe de recherche), les travaux menés à l'OLST gravitent autour d'une même approche générale, d'une façon à la fois souple (évolution des fondements théoriques et méthodologiques) et cohérente (convergence au niveau des buts de recherche et de la vision générale de la langue). Les deux grands axes de recherche de l'OLST — linguistique et terminologie-traduction — se renforcent mutuellement, chacun apportant à l'autre un éclairage original sur la façon de parvenir, notamment, à la construction de grandes bases de données lexicales. Après avoir pratiqué pendant cinq années une collaboration fondée avant tout sur des échanges d'idées, des partages de compétences, etc., les membres de l'OLST oeuvrent maintenant à concrétiser l'approche originale qu'ils ont graduellement développée sous la forme de nouveaux « produits de recherche ». Dans le cadre du présent programme, nous effectuons deux types de tâches. Premièrement, nous unifions nos descriptions lexicales — de langue générale et de langues de spécialité —, en élaborant des structures et des formats d'encodage qui les rendent entièrement compatibles. Il s'agit en fait de développer et mettre en application un métamodèle lexical, qui englobe les problématiques de la lexicographie et de la terminologie. Deuxièmement, nous cherchons à rendre compatibles nos méthodes de description, en particulier pour ce qui est du recours aux données linguistiques des corpus informatisés et des différentes techniques de construction des entrées de dictionnaires et des fiches de bases de données lexicales.

2. Projets et état d'avancement des travaux

Cette section décrit les projets spécifiques sur lesquels ont travaillé les chercheurs de l'OLST durant la période s'étendant de mai 2004 à mai 2006.

2.1 Dictionnaires et bases de données lexicales

Analytisches Handelslexikon / Dictionnaire analytique de la distribution / Analytical Dictionary of Retailing (responsable : Jeanne Dancette). Recherche de la terminologie allemande du *dictionnaire*; recherche des divergences conceptuelles dans les réalités décrites, recherche des contextes des termes allemands recensés; rédaction des définitions. Intégration des données dans la base de données générale (terminé). En cours : travail de mise en ligne de cette composante de langue allemande du *Dictionnaire*.

DiCo, base de données lexicale des dérivations sémantiques et collocations du français (responsables : Igor Mel'cuk et Alain Polguère). Ce projet a débuté il y a plus de dix ans et a atteint maintenant sa maturité. Dans le cadre des activités présentes de l'OLST, nous travaillons à (i) étendre la couverture du DiCo, (ii) améliorer sa cohérence et son homogénéité, (iii) développer des outils d'accès aux données lexicographiques et de modification de celles-ci.

DiCoInfo, Dictionnaire fondamental de l'informatique et de l'Internet (responsable : Marie-Claude L'Homme). En 2004-2006, le travail suivant a été fait sur le dictionnaire : précision de la nomenclature (env. 1800 articles), établissement de distinctions sémantiques; rédaction d'articles, mise en ligne d'articles dont la rédaction est terminée (plus de 600), enrichissement de la liste de termes reliés (plus de 20 000 liens sont décrits). En ce moment, un projet est en cours pour élaborer une version grand public du DiCoInfo.

Dictionnaire analytique de la distribution/Analytical Dictionary of Retailing (responsable : Jeanne Dancette). Mise en ligne du *dictionnaire* : toutes les données du dictionnaire papier (documents Word) ont été transférés sur fichiers Filemaker; les relations sémantiques établies entre les concepts clés et les vedettes du *Dictionnaire* ont été saisies et également converties en dossiers Filemaker. L'état d'avancement de la mise en ligne du dictionnaire progresse. La recherche par terme permet d'accéder aux fiches descriptives des 350 termes clés de langue anglaise.

Constitution d'une base de données permettant de gérer des fiches terminologiques et des textes multilingues de référence - domaine : sciences économiques et sociales (Responsable : Jeanne Dancette). La recherche est orientée autour des thèmes : *entreprises multinationales, relations du travail, mondialisation*. (travail en cours; recherche des meilleurs logiciels et adaptation aux besoins).

Mise au point d'une base de connaissances bilingues en biologie cellulaire et moléculaire (responsable : Sylvie Vandaele). Il s'agit de la prolongation d'un projet déjà entrepris. Le travail a consisté à améliorer la structure relationnelle d'une base de connaissances spécialisées, à trier les données déjà accumulées et à corriger les entrées

existantes (environ 500 entrées pour chacune des deux langues envisagées, l'anglais et le français).

Installation de l'interface Dicouèbe d'accès en ligne aux données du DiCo (responsables : Alain Polguère avec Jacques Steinlin et Sylvain Kahane de Paris 7). Collaboration avec une équipe du laboratoire du CNRS Lattice-Talana (Paris 7) en vue de l'écriture d'un compilateur de la base DiCo sous forme de tables SQL et de la construction d'une interface Web à ces tables. La programmation a été effectuée à Paris et, depuis l'automne 2004, nous avons mis en place un site Dicouèbe (= accès web aux données du DiCo) sur le site de l'OLST. L'adresse du Dicouèbe est <http://www.olst.umontreal.ca/dicouebe>.

Version coréenne du DiCoInfo (responsables : Hee Sook Bae et Marie-Claude L'Homme). La méthodologie mise au point pour le DiCoInfo français a été adaptée au coréen. Environ 100 articles sont en cours de rédaction et seront terminés d'ici juin 2006. Une fois terminés, ces articles seront placés dans le site de l'Observatoire de linguistique Sens-Texte.

2.2 Analyses de données lexicales

Analyse de collocations dans le domaine médical (responsable : Sylvie Vandaele). L'objectif du projet est de repérer les collocations importantes au sein des corpus de médecine. Deux volets sont en cours : le premier consiste à faire appel à un système d'annotation sémantique de corpus, de façon à en extraire l'information pertinente; le deuxième consiste à utiliser un système d'extraction fondé sur un étiquetage syntaxique (Syntex; système mis au point par Didier Bourigault, Toulouse – MDu, MDa).

Antonymes en corpus d'informatique (responsable : Marie-Claude L'Homme). Cette étude vise à décrire les liens d'antonymie exprimés dans un corpus d'informatique (elle s'inscrit dans un projet plus vaste qui cherche à mettre au point des méthodes de détection automatique de relations paradigmatiques, projet réalisé en collaboration avec Pascale Sébillot de l'Université de Rennes-1 et Alain Polguère). Nous avons extrait les paires de termes contraires apparaissant dans la même phrase et faisons actuellement le relevé des marqueurs linguistiques de la relation d'opposition dans une base de données.

Étude comparative des marqueurs de relation causale en corpus spécialisés (responsables Marie-Claude L'Homme et Elizabeth Marshman). Cette étude s'inscrit dans un axe de recherche qui consiste à repérer les contextes contenant une relation causale. Nous cherchons à déterminer si l'appel aux classes sémantiques des arguments d'un marqueur ambigu peut permettre de le désambiguïser. Une liste d'environ 35 marqueurs français a été constituée à partir d'un corpus médical et d'un corpus d'informatique; des contextes sont extraits de corpus médicaux et informatiques. L'analyse des contextes contenant un marqueur est en cours.

Étude des relations lexico-sémantiques (RLS) (Dancette 2003) dans le Dictionnaire analytique de la distribution /Analytical Dictionary of Retailing (responsable : Jeanne Dancette). Le modèle des RLS (Dancette 2003), comprenant 27 classes de RLS, a été testé et mis au point, puis systématiquement appliqué aux termes figurant dans le *Dictionnaire*.

2.3 Modélisation de données lexicales

Modélisation lexicographique des locutions (responsables : Alain Polguère et Sara-Anne Leblanc). Débuté en 2004, ce travail a permis la construction d'un ajout à la base DiCo d'une base complémentaire décrivant environ 800 locutions du français. La recherche se poursuit dans le cadre du travail de thèse de Sara-Anne Leblanc. Il s'agit notamment d'identifier les problèmes spécifiques posés par la modélisation lexicographique des locutions, et d'y remédier. La recherche touche directement l'interface entre lexique et grammaire.

Notions fondamentales de lexicologie, lexicographie et terminologie (responsable : Alain Polguère) : Il s'agit de construire un modèle ontologique des notions structurant les trois domaines en question. C'est un travail qui va s'étaler sur toute la durée du FQRSC et qui, jusqu'à présent, s'est alimenté des activités suivantes : production d'un rapport sur la structure des différentes bases lexicographiques et terminologiques de l'OLST, organisation par Marie-Claude L'Homme et Sylvie Vandaele d'un atelier sur les liens entre terminologie et lexicologie/lexicographie (avec publication subséquente d'un recueil), début de la construction d'une ontologie Protégé des notions de base. Ce dernier travail est temporairement en attente que d'autres travaux soient plus avancés (notamment celui sur la conceptualisation des systèmes lexicaux -- voir ci-dessous).

Système lexical du DiCo (responsable : Alain Polguère). Il s'agit de développer un nouveau type de structuration formelle des lexiques, appelée « système lexical ». Cette structure en graphe permet notamment une meilleure navigation dans les données lexicales et permet de modéliser des opérations de haut niveau --du type recherche, mise à jour ou inférence-- effectuées sur ce type de données. Le travail a commencé fin 2005 avec le début de la compilation des informations du DiCo sous forme de système lexical. Le transfert est presque complété en début 2006. La recherche doit se poursuivre sur la structure de données elle-même et sur son exploitation dans différents domaines (lexicologie, lexicographie, terminologie, linguistique appliquée...). Nous envisageons aussi l'écriture d'un éditeur lexicographique fonctionnant directement sur le système lexical du DiCo (pour remplacer notre méthode d'édition actuelle fondée sur l'édition de fiches lexicographiques "traditionnelles" au moyen d'un système de gestion de bases de données).

2.4 Corpus

Constitution d'un corpus étalon annoté pour l'acquisition automatique de termes (responsable : Patrick Drouin). Ce projet avait pour objectif l'élaboration d'un corpus étalon pouvant être utilisé pour évaluer les performances des logiciels d'acquisition automatique de termes. La terminologie contenue dans le corpus a été annotée manuellement, ce qui rend possible l'utilisation des mesures de précision et de rappel et la comparaison des résultats obtenus par les outils électroniques au jugement humain. Le corpus est constitué de documents tirés du Web et portant sur le domaine du tourisme, plus précisément de la réservation en ligne.

Constitution de corpus en médecine (responsable : Sylvie Vandaele, cofinancé par des fonds de l'Université de Montréal). Le projet avait pour but de constituer des corpus en anglais (env. 300 000 mots) et en français (env. 566 000 mots), ainsi que des corpus bilingues parallèles (env. 50 000 mots). Le projet a nécessité la mise au point de procédures informatiques spécifiques pour la conversion de fichiers en format PDF en format .txt exploitable par les concordanciers.

Constitution de corpus en sciences économiques et sociales (Responsable : Jeanne Dancette). Ce projet a mené à la constitution d'un corpus de textes de langues française et anglaise (3 millions de mots chacun) et de langue espagnole (2 millions de mots), ainsi qu'un corpus de textes parallèles bilingues et multilingues provenant des organisations internationales. Le projet a nécessité la numérisation de plusieurs ouvrages spécialisés, ainsi que la conversion des fichiers PDF trouvés sur Internet, en format.txt exploitables par les logiciels de recherche. Le travail consiste actuellement à rendre le corpus interrogeable par auteurs, titres, mots clés. La recherche des corpus est orientée autour des thèmes : entreprises multinationales, relations du travail, mondialisation.

2.5 Traitement automatique de la langue

Acquisition automatique de collocations nom-verbe (responsables : Vincent Claveau et Marie-Claude L'Homme). Nous avons mis au point et testé une méthode fondée sur la programmation logique inductive afin d'acquérir des collocations composées d'un verbe et d'un nom dans lesquelles les verbes ont un sens de réalisation. Des règles sont inférées à partir des contextes dans lesquels apparaissent les membres de couples valides, puis projetées sur le corpus afin d'acquérir de nouvelles paires.

Acquisition de couples de termes apparentés morphologiquement (responsables Vincent Claveau et Marie-Claude L'Homme). Nous avons mis au point et évalué une méthode visant à acquérir des paires de termes apparentés morphologiquement et à étiqueter le lien sémantique entre les termes. Cette méthode exploite l'analogie entre les termes afin de générer des règles. Ces règles sont ensuite appliquées à une liste de

termes spécifiques identifiés par TermoStat (mis au point par Patrick Drouin) pour retrouver de nouvelles paires.

Dépistage de la variation topolectale dans le domaine de l'automobile à partir de corpus spécialisés régionaux (responsable : Patrick Drouin). Dans la phase initiale de ce projet, nous avons constitué, à partir de sites Web spécialisés dans le domaine de l'automobile, des corpus électroniques régionaux pour divers pays francophones (Canada, France, Suisse et Belgique). Dans un premier temps, ces corpus feront l'objet de comparaison les uns par rapport aux autres afin de faire émerger les particularités lexicales de chacun de ces pays. La deuxième phase de ce projet, beaucoup plus ambitieuse, prévoit l'utilisation du moteur de recherche Google pour le dépistage de la variation topolectale pour les régions étudiées dans la première phase. Les résultats obtenus à partir du moteur de recherche seront présentés au sein d'une page Web qui comportera un module de veille permettant de visualiser les variations de fréquence dans le temps et l'espace des diverses variantes terminologiques régionales.

Identification des particularités lexicales régionales sur une base de comparaison de corpus journalistiques (responsable : Patrick Drouin). Ce projet visait l'application de techniques mises de l'avant en terminologie pour l'identification automatique de la terminologie dans les corpus spécialisés à des corpus journalistiques régionaux afin de tester la pertinence d'une telle approche pour le dépistage des usages lexicaux régionaux. Des expérimentations initiales ont été effectuées, ce qui a permis d'identifier des pistes de recherche intéressantes ainsi que des difficultés propres aux corpus non spécialisés devant être surmontées.

Impact lexical des événements de septembre 2001 sur la terminologie du terrorisme (responsable : Patrick Drouin). Cette étude vise à observer et décrire l'impact que les événements du 11 septembre 2001 ont eu sur la terminologie liée au domaine du terrorisme. L'étude a été réalisée sur un corpus spécialisé constitué de textes produits entre 1996 et 2005, ainsi que dans un corpus de presse plus large afin d'observer les variations de fréquences associées aux termes du domaine ainsi que les néologismes entourant ces événements.

Repérage de néologismes en biomédecine (responsables : Patrick Drouin et Sylvie Vandaele). L'objectif est de mettre au point une procédure de repérage des néologismes à partir de corpus constitués de résumés en anglais issus de Medline grâce à une approche mettant en jeu une comparaison de sous-corpus constitués de façon thématique.

3. Personnel

3.1 Professionnels de recherche

Stéfan Popovic, assistant de recherche pendant la première année du projet. Administration, élaboration du site web, implication dans les interfaces en lignes pour nos BD lexicales et terminologiques.

3.2 Étudiants

Dominic Anctil (maîtrise en linguistique) : construction de fiches lexicographiques, travail sur la hiérarchie d'étiquettes sémantiques du DiCo.

Julie Beaulieu (baccalauréat en traduction) : participation à la rédaction des articles du DiCoInfo.

Nancy Bourassa (maîtrise en traduction) : constitution d'un corpus étalon annoté pour l'acquisition automatique de termes.

Samia Bouchaddakh (doctorat en linguistique) : rédaction d'un document de synthèse sur les différentes BD lexicales et terminologiques en développement à l'OLST, mise au point d'une transcription d'un corpus oral.

Sylvie Boudreau (maîtrise en linguistique/doctorat en traduction) : constitution de corpus en médecine : mise au point des procédures informatiques.

Sébastien Cabot (baccalauréat en mathématique) : programmation d'un système d'analyse des définitions formelles de la base BDéf.

Mathilde Cloutier (maîtrise en traduction) : participation à la rédaction des articles du DiCoInfo.

Christian Collet (maîtrise en traduction) : conception de la base de données relationnelles du dictionnaire de la mondialisation, comprenant les banques de textes, les fiches et les références, ainsi que les informations pratiques sur les auteurs et le statut des fiches.

Mireille Daoust (Maîtrise en traduction) : analyse des collocations dans le domaine biomédical.

Louis-Philippe Dargis (baccalauréat en traduction) : Analyse des marqueurs de relation d'opposition dans un corpus d'informatique.

Maggie Dubé (baccalauréat en traduction) : mise au point d'une base de connaissances bilingues en biologie cellulaire et moléculaire; analyse des collocations dans le domaine biomédical.

Andréanne Germain-Rouleau (baccalauréat en linguistique) : construction de fiches lexicographiques, mise au point d'une transcription d'un corpus oral.

Anne-Laure Jousse (doctorat en linguistique) : construction de fiches lexicographiques, étude et affinement du système d'encodage des liens lexicaux au moyen des fonctions lexicales.

Sara-Anne Leblanc (baccalauréat en linguistique/doctorat en linguistique) : construction de fiches lexicographiques, modélisation des locutions dans le DiCo, mise au point d'une transcription d'un corpus oral; identification des particularités lexicales régionales sur une base de comparaison de corpus journalistiques.

Leslie Lubin (maîtrise en traduction) : mise au point d'une base de connaissances bilingues en biologie cellulaire et moléculaire; analyse de collocations dans le domaine médical : mise au point d'annotation sémantique de corpus.

Elizabeth Marshman (doctorat en traduction) : étude comparative des marqueurs de relation causale en corpus spécialisés; analyse de collocations dans le domaine médical : mise au point d'annotation sémantique de corpus.

Annie Paquin (maîtrise en linguistique) : impact lexical des événements de septembre 2001 sur la terminologie du terrorisme.

Johanne Rabeau (baccalauréat en traduction) : dépistage de la variation topolectale dans le domaine de l'automobile à partir de corpus spécialisés régionaux.

Sabine Schnittka (maîtrise en traduction) : élaboration de fiches de langue allemande sur *Analytisches Handelslexikon*.

Ophélie Tremblay (doctorat en linguistique) : mise au point de descriptions dans le champ sémantique des sentiments, recherche sur l'interfaçage des modélisations formelles dans un contexte d'enseignement du lexique.

3.3 Stagiaires

César Antonia Aguilar, Université nationale autonome de Mexico, Mexique (stagiaire de doctorat) : repérage de patrons définitoires dans un corpus d'informatique espagnol

Marie Ans, Université Stendhal, Grenoble, France (stagiaire de maîtrise) : Appariements sémantiques de termes sur corpus spécialisé, sans autres ressource que la structure lexicale et morphologique des termes.

Hee Sook Bae, Centre KorTerm, Séoul, Corée (stagiaire postdoctorale) : version coréenne du DiCoInfo.

Auxiliadora Barrios, Universidad Complutense, Madrid (stagiaire de doctorat) : application des principes de la lexicologie explicative et combinatoire à la construction de représentations type DiCo pour l'espagnol.

Sabrina Ben Hamou, Université Stendhal, Grenoble, France (stagiaire de maîtrise) : modélisation de la combinatoire syntactique des locutions dans la base de données lexicographiques DiCo.

Pétronille Bolle, Université Stendhal, Grenoble, France (stagiaire de maîtrise) : adaptation des structures actanciennes du DiCoInfo pour une version grand public.

Vincent Claveau, Université de Rennes, France (stagiaire postdoctoral) : acquisition automatique de collocations nom-verbe et acquisition de paires de termes apparentés morphologiquement.

Dominic Forest, Université du Québec à Montréal (stagiaire postdoctoral) : analyse thématique de documents.

Sylwia Ozdowska, Université de Toulouse-le-Mirail, France (stagiaire de doctorat) : acquisition de collocations dans des corpus comparables à partir des caractéristiques syntaxiques des collocatifs.

4. Autres activités

4.1 Séminaire, colloques

Les approches lexicographiques et terminologiques sont-elles compatibles ?, dans le cadre du 72^e Congrès de l'ACFAS 2004, colloque organisé par Marie-Claude L'Homme et Sylvie Vandaele

Séminaires OLST-RALI : Organisation de séminaires hebdomadaires en collaboration avec le RALI (Recherche appliquée en linguistique informatique, Département d'informatique et de recherche opérationnelle)

4.2 Professeurs, chercheurs invités

Françoise Boch (LIDILEM – Département de didactique Université Stendhal, Grenoble 3), août 2004.

Didier Bourigault (Équipe de Recherche en Syntaxe et Sémantique de Toulouse, Université Toulouse-le-Mirail), octobre 2004

Patrick Hanks (Gastwissenschaftler, Digitales Wörterbuch der Deutschen Sprache, Berlin-Brandenburgische Akademie der Wissenschaften, Berlin, Germany / Adjunct Professor, Department of Computer Science, Brandeis University, Waltham, MA, USA), mars 2006.

Stéphane Huet et Pascale Sébillot (Équipe TexMex, Irisa / Université de Rennes 1), juillet 2005

Pascale Sébillot (Équipe TexMex, Irisa / Université de Rennes 1), juillet 2004

Suzanne Simon-Alt (ATILF-CNRS, Nancy), mai 2005

4.3 Collaborations

Centre de recherche interfacultaire sur la mondialisation du travail (CRIMT) (G. Murray, M. Coutu).

Équipe TexMex, Irisa / Université de Rennes (Pascale Sébillot) : Détection automatique de relations paradigmatiques.

Philippe Langlais, RALI (Recherche appliquée en linguistique informatique), Université de Montréal.

Équipe de recherche en syntaxe et sémantique (Didier Bourigault) : utilisation et évaluation de Syntex, un analyseur syntaxique.

Convention d'échanges qui nous liait à l'Université Stendhal (Grenoble) (renouvellement), sous forme d'une convention tripartite incluant maintenant l'Université de Cologne.

5. Publications

a. Livres ou chapitres de livres

Iordanskaja, L. & I. Mel'čuk (2004). "The Meaning and Cooccurrence of Russian NEMNOGO 'A Little'", In Apresjan, Ju. D. (ed.), *Sokrovennye smysly: slovo, tekst, kul'tura. Sbornik statej v chest' N. D. Arutjunovoj*, Moskva: Jazyki slavjanskoj kul'tury, pp. 112-127.

L'Homme, M.C. (2005). "Lexical Semantics for Terminology. A Case Study: *Dictionary of Computing*", In Fuertes, P. (ed.). *Lengua y Sociedad*:

- Investigaciones recientes en Lingüística Aplicada*, Valladolid (Espagne): Universidad de Valladolid, pp. 233-245.
- L'Homme, M.-C. & E. Marshman (2006, sous presse). "Extracting terminological relationships from specialized corpora", In Bowker, L. (ed.). *Lexicography, Terminology, Translation: Text-Based Studies In Honour of Ingrid Meyer*. Ottawa: Ottawa University Press, 20 pages.
- L'Homme, M.C. & S. Vandaele (éd.) (à paraître). *Lexicographie et terminologie : compatibilité des modèles et des méthodes*, Ottawa : Presses de l'Université d'Ottawa.
- L'Homme, M.C. & S. Vandaele (à paraître). "Introduction", dans L'Homme, M.C. & S. Vandaele (éd.). *Lexicographie et terminologie : compatibilité des modèles et des méthodes*, Ottawa : Presses de l'Université d'Ottawa, 20 pages.
- Mel'čuk, I. (2004). "Zero Sign Revisited", In Vinogradov, V.A. (ed.), *Semiotika, poëtika, lingvistika. K stoletiju so dnja roždenija A. A. Reformatskogo*, Moskva: Jazyki slavjanskoj kul'tury, pp. 269-298.
- Mel'čuk, I. (2004). "Meaning in Nominal Cases", In Volodin, A. (ed.), *Tipologičeskie obosnovanija v grammatike. K 70-letiju professora V. S. Xrakovskogo*, Moskva: Znak, pp. 337.
- Mel'čuk, I. (2004). "Opredelenie kategorii zaloga i isčislenie vozmožnyx zalogov: 30 let spustja [La définition de voix et le calcul de voix possibles : 30 ans après]", In Xrakovskij, V., A. Mal'čukov & S. Dmitrenko (eds.), *40 let Sankt-Peterburgskoj tipologičeskoj škole*, Moskva: Znak, pp. 286-314.
- Mel'čuk, I. (2006). *Aspects of the Theory of Morphology* : Berlin/New York: Mouton de Gruyter.
- Polguère A. (à paraître). « *Soleil insoutenable et chaleur de plomb* : le statut linguistique des greffes collocationnelles », In L'Homme, M.C. & S. Vandaele (éd.). *Les approches lexicographiques et terminologiques sont-elles compatibles ?*, Ottawa : Presses de l'Université d'Ottawa.
- Polguère A. & O. Tremblay (à paraître). "Élaboration d'un 'site lexical' pour l'aide à la rédaction", In Grossmann, F. & S. Plane (éd.). *Lexique et production verbale : vers une meilleure intégration des apprentissages lexicaux*, Villeneuve-d'Ascq : Presses Universitaires du Septentrion.

b. Articles dans revues avec comité de lecture

- Claveau, V. & M.C. L'Homme (2006). "Discovering and Organizing Noun-Verb Collocations in Specialized Corpora Using Inductive Logic Programming", *International Journal of Corpus Linguistics* 11(2), pp. 73-107.

- Dancette, J. (2005). "Questions de conception ou de conceptualisation ? Du texte à la carte conceptuelle en traduction", *Le langage et l'homme* 40(1), pp. 107-122.
- Dancette, J. & M.C. L'Homme (2004). "Building Specialized Dictionaries Using Lexical Functions", *Linguistica Antverpiensia* 3, pp. 113-132.
- L'Homme, M.C. (2005). "Sur la notion de 'terme'", *Meta* 50(4), pp. 1112-1132.
- L'Homme, M.C. (2005). "Conception d'un dictionnaire fondamental de l'informatique et de l'Internet : sélection des entrées", *Le langage et l'homme* 40(1), pp. 137-154.
- Lemay, C., M.C. L'Homme & P. Drouin (2005). "Two Methods for Extracting Specific Single-word Terms from Specialized Corpora: Experimentation and Evaluation", *International Journal of Corpus Linguistics* 10(2), pp. 227-255.
- Mel'čuk, I. (2004). "Actants in Semantics and Syntax I: Actants in Semantics", *Linguistics* 42(1), pp. 1-66.
- Mel'čuk, I. (2004). "Actants in Semantics and Syntax II: Actants in Syntax", *Linguistics* 2(2), pp. 247-291.
- Mel'čuk, I. (2004). "Verbes supports sans peine", *Linguisticae Investigationes* 27(2), pp. 203-217.
- Mel'čuk, I. (2004). "La non-compositionnalité en morphologie linguistique", *Verbum* 26(4), pp. 439-458.
- Mel'čuk, I. & A. Polguère (à paraître). "Dérivations sémantiques et collocations dans le DiCo/LAF", *Langue française*, numéro spécial sur la collocation « Collocations, corpus, dictionnaires », sous la direction de P. Blumenthal et F. J. Hausmann.

c. Articles dans actes de congrès

- Anctil, D. (2005). "The notion of Lexical Error in the Framework of Explanatory Combinatorial Lexicology", In *Proceedings of the Second International Conference on the Meaning-Text Theory*, Moscow, pp. 9-19.
- Barque, L. (2006, à paraître). "Un lexique hybride pour le traitement de la polysémie", In *Actes du Colloque « Description linguistique pour le TAL »*, ACFAS 2006, Montréal.
- Barque, L. & A. Polguère (2005). "Application du métalangage définitionnel de la BDéf au traitement formel de la polysémie", In *Actes de TALN 2005*, Dourdan, pp. 391-397.

- Bourassa, N. & P. Drouin (2005). "Exploitation d'Internet dans le cadre du traitement assisté par ordinateur de la terminologie des domaines virtuels", In *Actes de la conférence META 50*.
- Carreño Cruz, S.I. (2005). "Análisis de la variación terminológica en corpus bilingüe especializado", In *IX Simposio Internacional de Comunicación Social. Actas*, Santiago de Cuba.
- Claveau, V. & M.C. L'Homme (2005). "Structuring Terminology by Analogy Machine Learning", *Terminology and Knowledge Engineering, TKE 2005*, Copenhagen Business School, Copenhagen (Danemark), 17-19 août 2005, pp. 301-312.
- Claveau, V. & M.C. L'Homme (2005). "Apprentissage par analogie pour la structuration de terminologie – Utilisation comparée de ressources endogènes et de ressources exogènes", In *Actes. Terminologie et intelligence artificielle, TIA 2005*, Université de Rouen, Rouen (France).
- Dancette, J. (2004). "Les représentations lexico-sémantiques (RLS), moyen de structuration des connaissances dans les domaines spécialisés", ISKO, Grenoble, pp. 30-37.
- Dancette, J. & S. Halimi (2004). "Repérage des relations lexico-sémantiques ; bilan d'une tentative de formalisation", In *Actes Terminologie, ontologie et représentation des connaissances*, Lyon III, pp. 30-38.
- Drouin, P. & H.S. Bae (2005). "Korean Term Extraction in the Medical Domain by Corpus Comparison", In *Terminology and Knowledge Engineering (TKE-2005)*, Copenhagen Business School, Copenhagen, pp. 349-361.
- Drouin, P. & P. Langlais (2006), "Évaluation du potentiel terminologique de candidats termes", In *Actes des 8^e Journées internationales d'analyse statistique des données textuelles (JADT-2006)*, Besançon, France, pp. 379-388.
- Drouin, P., A. Paquin & N. Ménard (2006), "Extraction semi-automatique des néologismes dans la terminologie du terrorisme", In *Actes des 8^e Journées internationales d'analyse statistique des données textuelles (JADT-2006)*, Besançon, France, pp. 389-400.
- Iordanskaja, L. & A. Polguère (2005). "Hooking up Syntagmatic Lexical Functions to Lexicographic Definitions", In *Proceedings of the Second International Conference on the Meaning Text Theory*, Moscow, pp. 176-186.
- L'Homme, M.C. (2004). "Sélection des termes dans un dictionnaire d'informatique : comparaison de corpus et critères lexico-sémantiques", *Euralex International Congress. Proceedings*. Lorient, France, pp. 583-593.
- L'Homme, M.C. (2004). "A Lexico-semantic Approach to the Structuring of Terminology", *Computerm 2004. Proceedings*, dans le cadre de Coling 2004, Genève (Suisse), pp. 7-14.

- Mel'čuk, I. (2005). "Meaning-Text Theory and the Conceptual Apparatus of Modern Linguistics: Towards a Better Description of Case and Voice in Maasai", In Apresjan, J. and L. Iomdin (eds.), *East West Encounter: Second International Conference on Meaning ⇔ Text Theory*, Moskva: Jazyki slavjanskoj kul'tury, pp. 285-303.
- Steinlin J., S. Kahane & A. Polguère (2005). "Compiling a 'Classical' Explanatory Combinatorial Lexicographic Description into a Relational Database", In *Proceedings of the Second International Conference on the Meaning Text Theory*, Moscou, pp. 477-485.
- Steinlin J., S. Kahane, A. Polguère & A. El Ghali (2004). "De l'article lexicographique à la modélisation objet du dictionnaire et des liens lexicaux", In *Proceedings of EURALEX'2004*, Lorient, pp. 177-186.

d. Communications non publiées

- Bae, H.S. & M.C. L'Homme (2006). "Converting a Monolingual Database into a Multilingual Specialized Dictionary", communication présentée lors du colloque *Multilingualism and Applied Comparative Linguistics (MACL-2006)*, Bruxelles, Belgique.
- Barque L. & A. Polguère (2004). "A Definitional Metalanguage for Explanatory Combinatorial Lexicography", Communication présentée au colloque *Word Structure and Lexical Systems: models and application*, Pavie, Italie.
- Carreño Cruz, S.I. (2004). "Analyse de la variation terminologique en corpus spécialisé anglais-espagnol", Communication présentée lors du Colloque *Les approches lexicographiques et terminologiques sont-elles compatibles ? 72^e Congrès de l'Association francophone pour le savoir (ACFAS)*, Université du Québec à Montréal, Québec.
- Carreño Cruz, S.I. (2005). "La variation terminologique en corpus parallèle et son incidence sur l'extraction de termes bilingues", Communication présentée lors du *Colloque international META 50^e anniversaire*. Université de Montréal, Canada.
- Dancette, J. (2004). "Concepts, Conceptualization and Comprehension in Translation" Communication présentée lors du Colloque 'Translation and conceptual maps', Pontifica Universidad catolica de Chile, Facultad de letras, Santiago, Chili.
- Dancette, J. (2004). "Des dictionnaires intelligents que la technologie rend accessibles aux traducteurs" Communication présentée lors de la conférence *Conference, Translating with Computer-Assisted Technology: Changes in Research, Teaching, and Practice*, Rome, Italie.
- Dancette, J., C. Collet & S. Dufour (2005). "Le système informatique du projet de Dictionnaire de la mondialisation du travail", Rapport de recherche présenté et

déposé au *CRIMT, Rethinking Institutions for Work and Employment*, Magog, Québec.

Leblanc, S.-A. (2005), “Modélisation de locutions dans une base de données DiCo”, communication présentée au congrès de l'ACFAS, Université du Québec à Chicoutimi, Québec.

Marshman, E. (2004). “L’exploitation de structures actanciennes et de classes sémantiques pour l’épuration des résultats de l’extraction semi-automatique de relations causales”, Communication présentée au colloque *Les approches terminologiques et lexicographiques, sont-elles compatibles ?*, dans le cadre du congrès ACFAS, Université du Québec à Montréal, Québec.

Racine, A. (2004). “Extraction automatique de néologismes dans les textes biomédicaux”, Communication présentée au colloque *Les approches terminologiques et lexicographiques, sont-elles compatibles ?*, dans le cadre du congrès ACFAS, Université du Québec à Montréal, Québec.

e. Autres

Dancette, J. (dir.) (2006). Le dictionnaire analytique de la distribution / Analytical Dictionary of Retailing, dictionnaire en ligne, avec exploitation des liens sémantiques entre les mots liés (<http://olst.ling.umontreal.ca/dad/>).

Drouin, P. (2004). TermoStat. Logiciel d'acquisition automatique de termes, Version Web. (http://olst.ling.umontreal.ca/~drouinp/termostat_web).

Drouin, P. et V. Claveau (dir.) (2006). Le Migou, concordancier en ligne permettant l'interrogation de corpus étiquetés généraux (ex. : *Le Monde*) et spécialisés (*médecine, informatique, mondialisation, etc.*). (<http://www.olst.umontreal.ca/migoufr.html>).

Jousse, A.-L. et A. Polguère (2005). *Le DiCo et sa version DiCouèbe. Document descriptif et manuel d'utilisation*, Département de linguistique et de traduction, Université de Montréal, 41 pages.

L’Homme, M.C. (dir.) (2006). DicoInfo. Dictionnaire fondamental de l’informatique et de l’Internet. Version Web (<http://olst.ling.umontreal.ca/dicoinfo>).

L’Homme, M.C. (ed.) (2006, forthcoming). Terminology 12(2), Special issue on the Processing of Terms in Dictionaries: New models and techniques.

Polguère, A., J. Steinlin & S. Kahane (2006). Dicouèbe. Interface d'accès en ligne aux données du DiCo (<http://www.olst.umontreal.ca/dicouebe>).

Popovic, S. et A. Polguère (responsables) (2006). Observatoire de linguistique Sens-Texte (OLST). Site Web du groupe de recherche (<http://www.olst.umontreal.ca/>).