

**Observatoire de linguistique Sens-Texte**

**OLST**

**Rapport d'activités**

**mai 2009 – avril 2010**

## Table des matières

|     |   |    |
|-----|---|----|
| 1.  | Membres réguliers et associés .....                   | 3  |
|     | Membres réguliers .....                               | 3  |
|     | Membres associés .....                                | 3  |
|     | Membres adjoints.....                                 | 3  |
| 2.  | Projets et état d'avancement des travaux.....         | 4  |
|     | Dictionnaires et bases de données lexicales.....      | 4  |
|     | Analyses de données lexicales .....                   | 5  |
|     | Modélisation des connaissances .....                  | 5  |
|     | Accès aux données lexicales .....                     | 5  |
|     | Corpus .....  | 6  |
|     | Traitement automatique de la langue .....             | 6  |
|     | Paraphrase .....                                      | 7  |
|     | Didactique du lexique .....                           | 7  |
| 3.  | Encadrement d'étudiants aux études supérieures .....  | 8  |
| 4.  | Personnel.....  | 11 |
|     | Étudiants .....                                       | 11 |
|     | Stagiaires.....                                       | 12 |
| 5.  | Autres activités .....                                | 13 |
|     | 5.1 Séminaires, colloques.....                        | 13 |
|     | 5.2 Professeurs, chercheurs invités .....             | 13 |
|     | 5.3 Collaborations .....                              | 13 |
| 6.  | Publications.....                                     | 15 |
| 6.1 | Livres ou chapitres de livres .....                   | 15 |
|     | 6.2 Articles dans revues avec comité de lecture ..... | 16 |
|     | 6.3 Articles dans des actes de congrès .....          | 18 |
|     | 6.4 Communications non publiées .....                 | 20 |
|     | 6.5 Autres .....                                      | 23 |

## **1. Membres réguliers et associés**

### **Membres réguliers**

Lyne Da Sylva (depuis le 5 février 2010)  
Patrick Drouin (directeur adjoint)  
Lidjia Iordanskaja  
Pascale Lefrançois  
Marie-Claude L'Homme (directrice)  
Elizabeth Marshman  
Igor Mel'čuk  
Jasmina Milićević  
Etienne Nguyen (jusqu'au 8 septembre 2009)  
Alain Polguère

### **Membres associés**

Amparo Alcina (TecnoLeTTra - Universitat Jaume I)  
Ophélie Tremblay (Université de Montréal, depuis le 15 février 2010)

### **Membres adjoints**

Benoît Robichaud

## **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 2009 à avril 2010.

### **Dictionnaires et bases de données lexicales**

**Definiens** (responsables : Lucie Barque – Lille 3 – et Alain Polguère). Ce projet a débuté à l'automne 2008 (sous la désignation de *Structuration et balisage sémantique des définitions du « Trésor de la Langue Française informatisé »*), durant le stage post-doctoral de Lucie Barque. Il vise à fournir une structuration explicite et formalisée des définitions du Trésor de la Langue Française informatisé (TLFi), sous forme de base de données XML de définitions lexicographiques. Pour cela, il s'appuie notamment sur les acquis du projet BDéf de construction d'une base formelle de définitions, dérivée du Dictionnaire explicatif et combinatoire du français contemporain (DECFC). La méthodologie de balisage des définitions et les outils d'édition XML ont tout d'abord été mis en place (avec rédaction par les deux coresponsables d'un guide destiné aux annotateurs). Le travail de balisage proprement dit a ensuite débuté, à partir des données source du TLFi, fournies par Jean-Marie Pierrel, directeur de l'ATLIF (CNRS, Nancy). Le projet est maintenant hébergé, sous le nom de Definiens, par l'ATILF, dans le cadre des projets internes du laboratoire, en collaboration avec l'OLST. Dans le cadre de cette collaboration, une partie des annotations est effectuée par une équipe d'étudiants de l'OLST, en complément des annotations effectuées à Nancy. Une première série de fichiers de définitions annotées sera mis en ligne par l'ATILF, sur le site du Centre National de Ressources Textuelles et Lexicales (CNRTL), cet été. Un article sur le projet a déjà été présenté en juin 2009 par les coresponsables au colloque MTT'09 de Montréal. Une autre publication, écrite conjointement avec Alexis Nasr (Marseille 2), a été acceptée pour EURALEX 2010 et sera présentée en juillet.

**DiCoEnviro** (responsable : Marie-Claude L'Homme). Ce projet vise à construire une base de données terminologique portant sur le domaine de l'environnement. La base de données sera élaborée selon les principes mis au point dans le cadre du DiCoInfo. Des corpus comportant des textes anglais en français et portant sur la thématique du réchauffement climatique ont été assemblés. Les termes français et anglais ont été extraits à l'aide de TermoStat et font actuellement l'objet d'une sélection. Quelque cent fiches sont actuellement rédigées. La mise en ligne des premiers articles est prévue pour la fin d'avril 2010.

**Version multilingue du DiCoInfo** (responsable : Marie-Claude L'Homme) : Des versions anglaise et espagnole du DiCoInfo sont en cours de rédaction. En anglais, près de 600 acceptions ont été distinguées et plus de 4000 liens sémantiques répertoriés. En outre, les liens d'équivalence avec le DiCoInfo français sont établis pour environ 900 lexies. En espagnol (responsables : Amparo Alcina et Marie-Claude L'Homme), la sélection des termes est amorcée par l'équipe de Castellón (un financement a été obtenu auprès de l'Université Jaume I pour ces travaux). Un séminaire de formation donné à

Castellón a permis de commencer la rédaction de quelques articles. Quelques articles sont en ligne depuis mars 2010.

## **Analyses de données lexicales**

*Annotation syntaxico-sémantique de contextes spécialisés* (responsable : Marie-Claude L’Homme et Guy Lapalme). Ce projet (financé par le Conseil de recherches en sciences humaines, CRSH) vise à fournir, pour des termes du DiCoInfo et du DiCoEnviro, une annotation syntaxico-sémantique. L’annotation porte sur les termes de nature prédicative ainsi que sur leurs actants et circonstants. L’équipe a mis au point un protocole d’annotation réalisé dans une structure XML. Jusqu’à présent, plus de 200 verbes du DiCoInfo et environ 50 verbes du DiCoEnviro sont annotés. Les premières annotations ont porté sur les verbes français. Toutefois, l’annotation des contextes verbaux anglais a commencé et celle des verbes espagnols débutera au cours de l’été 2010. Quelques noms de nature prédicative sont également en cours d’annotation.

*Étude du lexique scientifique transdisciplinaire* (responsable Patrick Drouin) : Cette étude s’inscrit dans un axe de recherche portant sur le lexique commun à divers domaines scientifiques. Le lexique scientifique transdisciplinaire (LST) fait l’objet de descriptions lexicographiques en anglais et en français. L’objectif de ces descriptions est double : identifier les sens des unités lexicales recensés en contexte et décrire les collocations au sein desquelles ces unités interviennent. Un corpus de 8 millions de mots en (4 millions en anglais et 4 millions en français) a été constitué et analysé. Le vocabulaire transdisciplinaire a été isolé dans les deux langues et une première description des sens est presque terminée.

## **Modélisation des connaissances**

*Gros Tas de Notions* (GTN) (responsable : Alain Polguère). Continuation du travail décrit dans les rapports précédents sur l’ontologie des savoirs lexicographiques et terminologiques. Le développement de l’ontologie des notions utilisées en lexicologie/lexicographie et terminologie s’est poursuivi avec la collaboration d’Ophélie Tremblay — étudiante de doctorat du Département de didactique jusqu’à sa soutenance de thèse à l’automne 2009. Une première mise en ligne expérimentale de cette ontologie a été effectuée mais n’est pas encore rendue publique (<http://olst.ling.umontreal.ca/gtn>). Le GTN a été exploité comme outil de structuration de stratégies pédagogiques d’enseignement des connaissances lexicales dans le cadre de la thèse de doctorat d’Ophélie Tremblay, soutenue à l’automne 2009 au département de didactique. Le GTN est utilisé comme base de connaissances dans le cadre du projet Lexitation (ontologisation des savoirs lexicographiques) d’Alain Polguère, financé par le CRSH.

## **Accès aux données lexicales**

**Accès aux collocations multilingues spécialisées** (responsables : Marie-Claude L’Homme et Benoît Robichaud en collaboration avec Patrick Leroyer). Ce projet consiste à permettre d’accéder aux traductions des collocations répertoriées dans le DiCoInfo et le

DiCoEnviro. Cet accès est réalisé en repérant les fonctions lexicales (FL) associées aux collocations décrites dans chacune des langues.

**Agrégation des contenus des bases de données lexicales de l'OLST** (responsable : Benoît Robichaud) : Ce site web est dédié à l'interrogation et à l'agrégation des contenus des bases de données lexicales électroniques de l'OLST. Pour l'instant, l'interface permet l'accès à quatre bases de données hébergées à l'OLST et extrait, pour les résultats correspondant à chaque requête, les unités lexicales, un contexte et un pointeur vers la fiche complète.

**Conception d'une interface multilingue au DiCoInfo et au DiCoEnviro** (responsables : Marie-Claude L'Homme et Benoît Robichaud). Ce projet consiste à adapter l'interface – d'abord conçue en français – à l'anglais et à l'espagnol. En outre, le module de recherche permet désormais l'extraction des équivalents des termes lorsqu'ils sont en ligne.

## **Corpus**

**Constitution d'un corpus de bitextes de vulgarisation dans le domaine de la santé** (responsable : Elizabeth Marshman). Ce projet vise la constitution d'un corpus de bitextes alignés anglais-français. Le corpus renferme des textes de vulgarisation tirés de sites Internet d'organismes canadiens dans le domaine de la santé. Le corpus compte actuellement environ 250 000 mots dans chaque langue, mais continue d'être alimenté.

## **Traitement automatique de la langue**

**Annotation automatique des actants** (responsables : Guy Lapalme et Marie-Claude L'Homme). Ce projet (financé par le Conseil de recherches en sciences humaines, CRSH) vise à alléger le travail d'annotation syntaxico-sémantique des termes français du DiConfo et du DiCoEnviro (voir ce projet décrit plus haut). La méthode comporte les étapes suivantes : 1) identification des participants; 2) distinction des actants et des circonstants; 3) attribution automatique de rôles sémantiques. Des expérimentations ont été réalisées pour la première tâche. Les résultats sont prometteurs. Une extension des modèles d'annotation à l'anglais et à l'espagnol doit commencer au cours de l'été 2010.

**Lexique transdisciplinaire** (responsable : Patrick Drouin) : L'objectif du projet est de repérer automatiquement le lexique scientifique transdisciplinaire (LST) au sein d'un corpus de thèses et d'articles scientifiques reliés à des domaines variés. Afin d'atteindre cet objectif, diverses méthodes de traitement automatique de la langue mettant faisant appel à des traitements linguistiques et statistiques sont mises de l'avant. Le corpus utilisé pour ce projet totalise 8 millions de mots, dont 4 millions en français et 4 millions en anglais. Dans chacune des langues, le corpus est constitué de 2 millions de mots tirés de thèses et de 2 millions de mots qui proviennent d'articles scientifiques.

**Texto4Science** (responsable : Patrick Drouin en collaboration avec Philippe Langlais du RALI) : Le projet Texto4Science a pour objectif de mettre en place une base de données de textos (courts messages échangés par téléphone cellulaire) en anglais et en français. Le projet s'inscrit dans le cadre du projet international SMS4science, initiative du Centre de Traitement Automatique du Langage de l'Université catholique de Louvain, en Belgique. Les textos reçus sont transcrits et annotés.

## **Paraphrase**

**Extraction de règles de paraphrasage à partir de corpus de textes parallèles bilingues** (responsable : Jasmina Milićević). Subventionné par le *Research Development Fund for the Humanities and Social Sciences* de Dalhousie University jusqu'en juin 2011, ce projet vise à repérer les liens paraphrastiques dans un corpus de textes littéraires bilingues (les paires de langues impliquées sont russe-français, français-anglais et espagnol-français.) La description des liens se fait sous la forme de règles de paraphrasage utilisables pour diverses tâches de TAL (reformulation, création des résumés, traduction automatique, etc.).

## **Didactique du lexique**

**Élaboration et validation d'un instrument d'évaluation du vocabulaire d'élèves québécois de 3e secondaire** (responsable: Pascale Lefrançois, en collaboration avec Marie-Éva de Villers). Ce projet consiste à développer un questionnaire évaluant différents aspects de la connaissance lexicale des élèves québécois francophones de 3e secondaire, à partir d'une variété de tâches fonctionnelles et métalinguistiques. Le questionnaire a été validé par des experts, puis par des passations expérimentales auprès d'élèves de trois régions du Québec. Une fois le questionnaire validé, il pourra être administré dans une enquête visant à faire le portrait des connaissances lexicales des jeunes Québécois.

### **3. Encadrement d'étudiants aux études supérieures**

**Marta Allard** (3e cycle, traduction, Université d'Ottawa). *Terminology Management for Translators in Translation Environments*. (codirection : Lynne Bowker et Elizabeth Marshman; en cours).

**Dominic Anctil** (3<sup>e</sup> cycle, didactique). *L'erreur lexicale au secondaire : analyse d'erreurs lexicales d'élèves de troisième secondaire en rédaction et description du rapport à l'erreur lexicale d'enseignants de français* (codirection : Pascale Lefrançois et Françoise Boch; en cours).

**Évelyne Arbour-Barbaud** (2<sup>e</sup> cycle, linguistique). *Dépendances syntaxiques du français: problèmes d'identification et d'encodage lexicographique* (Alain Polguère; en cours).

**Baris Bilgen** (2e cycle, traduction, Université d'Ottawa). *Investigating Terminology Management for Conference Interpreters* (codirection: Elizabeth Marshman et Annie Brisset; soutenu en juillet 2009).

**Samia Bouchaddakh** (3e cycle, linguistique). Application des principes de la lexicologie explicative et combinatoire à la construction d'un dictionnaire bilingue informatisé français-arabe (Gaétane Dostie Université de Sherbrooke – dir. et Alain Polguère codir.; thèse déposée, soutenance convoquée pour juin 2010).

**Sahara Iveth Carreño Cruz** (3<sup>e</sup> cycle, traduction). *Méthodes de regroupement de variantes terminologiques en extraction automatique de termes : application à des corpus spécialisés anglais, espagnol et français* (codirection : Marie-Claude L'Homme et Patrick Drouin; en cours).

**Sophie Comeau** (2<sup>e</sup> cycle, linguistique). *Partager le savoir du lexicographe : extraction et modélisation ontologique des savoirs lexicographiques* (codirection : Alain Polguère et Pascale Lefrançois; terminé en janvier 2010).

**Widade El Cherif** (2<sup>e</sup> cycle, linguistique). *Patrons de dérivation de noms d'agent en français. Perspectives synchronique et diachronique* (Jasmina Milićević; en cours.).

**Marie-Claude Demers** (2<sup>e</sup> cycle, traduction). *Méthode d'enrichissement de la nomenclature d'un dictionnaire général : ajout de termes d'informatique* (Marie-Claude L'Homme; en cours).

**Stephanie Doyle** (2e cycle, linguistique). *Apprentissage et enseignement des catégories flexionnelles verbales du français et de l'anglais : étude comparative*. (Jasmina Milićević; en cours.)

**Fadila Hadouche** (3<sup>e</sup> cycle, informatique). *Annotation automatique de rôles actanciels* (codirection : Guy Lapalme et Marie-Claude L'Homme; en cours).



**Joshua Holden** (3<sup>e</sup> cycle, linguistique). *Description du lexique fondamental du déné suliné en relation avec la morphologie du verbe déné* (codirection : Igor Mel'cuk et Alain Polguère; en cours).

**Anne-Laure Jousse** (3<sup>e</sup> cycle, linguistique). *Modèle de structuration des relations lexicales fondé sur le formalisme des fonctions lexicales* (cotutelle : Sylvain Kahane – Paris 10 & Paris 7 (pour l'inscription de la thèse) – et Alain Polguère; thèse déposée, soutenance prévue pour juillet 2010).

**Daniel Alberto Labonia** (3<sup>e</sup> cycle, linguistique). *Modélisation sémantique des ressources linguistiques en extraction automatique d'information* (Marie-Claude L'Homme; en cours).

**Kim Lacroix** (2e cycle, traduction, Université d'Ottawa). *Le parachutisme sportif au Canada : un domaine évolutif en quête d'une terminologie propre. Fichier terminologique bilingue commenté* (Elizabeth Marshman; soumis en avril 2010).

**Audrey Laroche** (3e cycle, linguistique). *Identification automatique de la polysémie à partir des profils de cooccurrence large des signifiants lexicaux* (codirection : Patrick Drouin et Alain Polguère)

**Annaïch Le Serrec** (3<sup>e</sup> cycle, traduction). *Analyse comparative de l'équivalence terminologique en corpus parallèle et en corpus comparable : application au domaine du changement climatique* (Marie-Claude L'Homme; en cours).

**Sara-Anne Leblanc** (3<sup>e</sup> cycle, linguistique). *Combinatoire des locutions françaises dans une perspective lexicographique : à l'interface entre lexique et grammaire* (codirection : Alain Polguère et Sylvain Kahane – Paris 10; en cours).

**Patricia Lee Men Chin** (3e cycle, linguistique appliquée). *Étude du rôle des logiciels d'ALAO dans le développement des stratégies d'apprentissage : une méta-analyse.* (codirection : Marie-Josée Hamel, Université d'Ottawa, et Jasmina Milićević; en cours.)

**François Louis** (3e cycle, linguistique). *Le vocabulaire d'un bretonnant de naissance de Pluméliau, région de Pontivy.* (codirection : Francis Favreau, Université de Rennes-2 et Jasmina Milićević; en cours.)

**Cheryl McBride** (2e cycle, traduction, Université d'Ottawa). *Translation Memory Systems: An Analysis of Translators' Attitudes and Opinions* (codirection : Lynne Bowker et Elizabeth Marshman; soutenu en septembre 2009).

**Janine Pimentel** (3<sup>e</sup> cycle, traduction). *Criteria for the Validation of Specialized Verb Equivalents: Applications in Bilingual Terminography* (codirection : Marie-Claude L'Homme et Rute Costa, Université nouvelle de Lisbonne; en cours).

**Caroline Poudrier (2e cycle).** *Évaluation de deux méthodes d'identification des liens lexicaux syntagmatiques : méthodes manuelle et statistique* (codirection : Alain Polguère et Patrick Drouin).

**Karine Pouliot (2<sup>e</sup> cycle, didactique).** *Élaboration d'un vocabulaire du lexique transdisciplinaire scientifique : principaux termes et collocations* (codirection : Pascale Lefrançois et Monique Cormier; en cours)

**Mélissa Samson-Fauteux (2e cycle, linguistique).** *Représentations du sens dans la théorie linguistique Sens-Texte.* (codirection : Igor Mel'čuk et Jasmina Milićević; en cours.)

**Ophélie Tremblay (3<sup>e</sup> cycle),** *Création d'une ontologie des connaissances métalexicales pour l'élaboration d'un module de cours en didactique du lexique destiné aux futurs maîtres au primaire en français langue maternelle* (codirection : Pascale Lefrançois et Jacqueline Bourdeau; soutenue en octobre 2009).

**Alexandra Tsedryk (3e cycle, linguistique).** *Compétence paraphrastique en français langue seconde.* (Jasmina Milićević; en cours.)

**David Wilton (2e cycle, linguistique).** *Description lexicographique des lexies dénotant des animaux dans un dictionnaire explicatif et combinatoire.* (Jasmina Milićević; en cours.)

**Ibtissam Yacoub (2<sup>e</sup> cycle, traduction).** *Recherche terminologique dans le domaine de l'environnement.* (Marie-Claude L'Homme; en cours)

## 4. Personnel

### Étudiants

**Ariane Bouchard** (Bacc. en traduction). Rédaction de fiches dans le DiCoInfo anglais.

**Gabriel Bernier Colborne** (baccalauréat en traduction). Constitution et annotation d'un corpus étalon pour l'acquisition automatique de termes dans le domaine de l'automobile.

**Sophie Comeau** (maîtrise en linguistique). Ontologisation de savoirs lexicographiques à partir de l'analyse d'enregistrements de sessions d'activités lexicographiques effectuées par des experts.

**Marie-Claude Demers** (maîtrise en traduction). Rédaction de fiches dans le DiCoInfo anglais.

**Suzanne DesGroseilliers** (Bacc. en traduction). Rédaction de fiches dans le DiCoEnviro.

**Claudia Fecteau** (Bacc. en linguistique). Balisage XML de définitions du TLFi pour l'identification des composantes définitionnelles – projet *Definiens*.

**Anne-Laure Jousse** (doctorat en linguistique). Modèle de structuration dynamique d'un lexique dans une perspective de transformation et de consultation.

**Marie-Ève Laneville** (maîtrise en traduction). Annotations contextuelles du DiCoInfo français; rédaction de fiches dans le DiCoEnviro.

**Audrey Laroche** (doctorat en linguistique). Programmation en vue de l'évolution du logiciel TermoStat (version 3.0).

**Thi Hoa Lê** (doctorat en linguistique soutenu). Désambiguïsation des étiquettes sémantiques du DiCo par rattachement aux acceptions correspondantes des vocables du TLFi.

**Annaïch Le Serrec** (doctorat en traduction). Annotations contextuelles et DiCoEnviro.

**Sara-Anne Leblanc** (doctorat en linguistique). Modélisation lexicographique des unités lexicales phraséologiques.

**Stéphanie Legros** (Bacc. en linguistique). Ontologisation de savoirs lexicographiques à partir de l'analyse d'enregistrements de sessions d'activités lexicographiques effectuées par des experts.

**Marie-Ève Paquette** (Bacc. en traduction). Rédaction de fiches dans le DiCoInfo anglais.

**Amélie Paulus** (baccalauréat en traduction). Constitution d'un corpus bilingue d'articles scientifiques en vue du repérage automatique du vocabulaire scientifique transdisciplinaire; descriptions lexicographiques du vocabulaire scientifique transdisciplinaire; annotation et transcription du corpus de textos pour le projet Textto4Science.

**Caroline Poudrier** (maîtrise en linguistique). Balisage XML de définitions du TLFi pour l'identification des composantes définitionnelles – projet *Definiens*.

**Janine Pimentel** (doctorat en traduction). Annotations contextuelles du DiCoInfo anglais.

**Eugénie Rompré Brodeur** (baccalauréat en traduction). Annotation et transcription du corpus de textos pour le projet Textto4Science.

**Mélissa Samson-Fauteux** (maîtrise en linguistique). Balisage XML de définitions du TLFi pour l'identification des composantes définitionnelles – projet *Definiens*.

**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.

## **Stagiaires**

**Geun-Seok Lim**, Seoul University (stagiaire post-doctoral). Description formelle de quelques éléments de base du lexique du coréen : le verbe *I-da* ≈ 'être'.

**Sébastien Marengo**, Université de Strasbourg 2 (stagiaire post-doctoral). Étude des structures régies par les gouverneurs syntaxiques adjectivaux en français.

**Aurélie Picton**, Université Toulouse-le-Mirail (stagiaire post-doctorale). Identification d'évolutions sémantiques dans la terminologie de domaines de spécialité.

**Esperanza Valero Domenech**, Universitat Jaume I (stagiaire doctorale), rédaction de fiches dans la version espagnole du DiCoInfo; traduction du manuel de l'utilisateur.

## **5. Autres activités**

### **5.1 Séminaires, colloques**

Préparation d'une table ronde « Différentes perspectives sur le traitement automatique des documents: l'indexation », dans le cadre du Congrès des milieux documentaires du Québec, Montréal, novembre 2010. (acceptée)

Préparation de *la 17<sup>e</sup> Conférence sur le Traitement automatique des langues naturelles* (TALN 2010), juillet 2010.

Préparation des *Journées lexicographie / terminologie de l'Observatoire de linguistique Sens-Texte* (OLST), juin 2010.

Préparation du *6e Défi Fouille de texte* (DEFT 2010), 23 juillet 2010, Université de Montréal, atelier dans le cadre de TALN.

Préparation du colloque « Approches sémiotiques au design », 10-11 mai 2010 dans le cadre de l'ACFAS, Université de Montréal.

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

### **5.2 Professeurs, chercheurs invités**

**Amparo Alcina** (Université Jaume I), juillet 2009

**Rute Costa** (Université nouvelle de Lisbonne), février 2010

### **5.3 Collaborations**

Centre de recherches en traduction et terminologie, Université Lyon-2, France (contact : Pascaline Dury). Travaux sur la nécrologie terminologique

Équipe de recherche en syntaxe et sémantique du Laboratoire CLLE-ERSS, Université Toulouse – le Mirail, France (contact : Didier Bourigault). Utilisation et évaluation de l'analyseur syntaxique *Syntex*.

Équipe Tecnolettra / Université Jaume I (contact : Amparo Alcina). Version espagnole du DiCoInfo.

Laboratoire ATILF CNRS (Nancy). Projet *Definiens* (voir description plus haut).

Laboratoire Centlex, Université d'Aarhus, Danemark (contact : Patrick Leroyer). Projet d'accès aux traductions multilingues spécialisées.

Laboratoire LICEF de la Télé-Université à Montréal (contact : Jacqueline Bourdeau).  
Projet Lexitation d'ontologisation des savoirs lexicographiques en vue de leur utilisation dans les disciplines de linguistique appliquée.

Laboratoire RALI de l'Université de Montréal (contact : Guy Lapalme). Conversion au format XML du DiCoInfo et prises en compte de données multilingues.

Université de Cologne (Italie) / Université Stendhal (France). Convention d'échanges tripartite pour une période de 3 ans entre ces universités et l'Université de Montréal.

## 6 Publications

### 6.1 Livres ou chapitres de livres

- Alcina, A. (2009) « Metodología y tecnologías para la elaboración de diccionarios terminológicos onomasiológicos ». In Alcina, A., E. Valero and E. Rambla (eds.) *Terminología y Sociedad del conocimiento*, Peter Lang, Berna. pp 33-58.
- Alcina, A. et M.C. L'Homme (eds.) (2009) *Proceedings of the First International Workshop on Terminology and Lexical Semantics (TLS'09)*. Observatoire de linguistique Sens-Texte (OLST), Montréal: Université de Montréal.
- Ancil, D. (2009) « Qu'est-ce qu'une erreur lexicale? ». In *Les Langues Modernes Revue de l'Association des Professeurs de Langues Vivantes*, vol. 103, no 1. Enseigner et apprendre le lexique.  
Version Web : [www.aplv-languesmodernes.org](http://www.aplv-languesmodernes.org)
- Beck, D., Gerdes, K., Milićević, J. et Polguère, A. (eds.) (2009) *Proceedings of the Fourth International Conference on Meaning-Text Theory (MTT'09)*. Observatoire de linguistique Sens-Texte (OLST), Montréal : Université de Montréal, 398 pages.
- Barrón-Cedeño, A., S. Gerardo; D. Patrick et S. Ananiadou (2009) « An Improved Automatic Term Recognition Method for Spanish », In *Computational Linguistics and Intelligent Text Processing*, vol. 5449, Berlin/Heidelberg, Springer, p. 125-136.
- Hamel, M.-J., J. Milićević et R. Mopoho (eds) (2009) *Proceedings of the 32nd Annual Meeting of the Atlantic Provinces Linguistics Association (PAMAPLA 32)*. Dalhousie University, Halifax, 14-15 novembre 2008.
- Iordanskaja, L. et I. Mel'čuk (2009) « Connotation (in Linguistic Semantics) ». In: S. Kempgen, P. Kosta, T. Berger & K. Gutschmidt (eds.) *The Slavic Languages. An International Handbook of their Structure, their History and their Investigation*, 2009, Berlin/New York: Mouton de Gruyter, 875-882.
- Iordanskaja, L. et I. Mel'čuk (2009) « Semantics of the Russian Conjunction POKA 'while, before, until' ». In: T. Berger, M. Giger, S. Kurt & I. Mendoza (eds), *Von grammatischen Kategorien und sprachlichen Weltbildern – Slavia von der Sprachgeschichte bis zur Politsprache. Festschrift für Daniel Weiss zum 60. Geburtstag*, 2009, München – Wien, 233-262 [= WSA, Sonderband 73].
- L'Homme, M.C. (2009) « A Methodology for Describing Collocations in a Specialized Dictionary ». In Nielsen, S. and S. Tarp (eds.). *Lexicography in the 21<sup>st</sup> Century. In Honour of Henning Bergenholtz*. Amsterdam/Philadelphia: John Benjamins, pp. 237-256.

- L'Homme, M.C. (2010) « Designing Terminological Dictionaries for Learners based on Lexical Semantics: The representation of actants ». In Fuertes-Olivera, P. (ed.). *Specialised Dictionaries for Learners*. Berlin/New York: Walter de Gruyter.
- L'Homme, M.C. (2010, à paraître) « Large Terminological Databases », In *International Encyclopedia of Lexicography*.
- Marshman, E. (soumis) « Challenges of Lexical Knowledge Patterns for Bilingual Terminology Work. » In L. Bowker (ed), *University of Ottawa School of Translation and Interpretation 35th Anniversary Volume*.
- Mel'čuk, I. et A. Polguère (2009) « Foreword ». In Polguère A. & Mel'èuk I. (eds.) *Dependency in Linguistic Description.*, pp. xiii--xxii.
- Polguère, A. et I. Mel'čuk (eds.) (2009) *Dependency in Linguistic Description*. Language Companion Series 111, John Benjamins, Amsterdam/Philadelphie, xxii–281 pages.

## 6.2 Articles dans revues avec comité de lecture

- Bowker, L. et E. Marshman. « Better integration for better preparation: Bringing terminology and technology more fully into translator training using the CERTT approach ». In *Terminology* 15(1): 60-87, 2009
- Drouin, P. (à paraître) « Extracting a bilingual transdisciplinary scientific lexicon ». In *Proceedings of eLexicography*, 11 pages.
- Drouin, P. et F. Doll (à paraître) « Potentiel terminologique, quel sens prendre ? ». In *Actes des 8es Journées scientifiques du Réseau LTT*, 11 pages.
- Dury, P. et P. Drouin (à paraître), « L'obsolescence des termes en langues de spécialité, une étude semi-automatique de la « nécrologie » en corpus informatisés, appliquée au domaine de l'écologie », In *Proceedings of the XVII European Symposium On Languages For Specific Purposes*, 15 pages.
- Hadouche, F., M.C. L'Homme & G. Lapalme (2010, à paraître) « Automatic Annotation of Actants », In *Les Cahiers du Cental*.
- Maroto, N. et A. Alcina (2009) « Formal description of conceptual relationships with a view to implementing them in the ontology editor Protégé », In *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 15 (2), 232-257.
- Iordanskaja, L. et I. Mel'čuk (2010; à paraître) « La coordination des adjectifs modificateurs en russe et en français (la conjonction russe I vs la conjonction française ET) ». In *Méta*



- L'Homme, M.C. (2010, à paraître) « Designing Specialized Dictionaries with Natural Language Processing. Examples of applications in the fields of computing and climate change », In *Les Cahiers du Cental*.
- L'Homme, M.C. et P. Leroyer (2009) « Combining the semantics of collocations with situation-driven search paths in specialized dictionaries », In *Terminology* 15(2), pp. 258-283.
- Le Serrec, A., M.C. L'Homme, P. Drouin et O. Kraif (2010) « Automating the compilation of specialized dictionaries: use and analysis of term extraction and lexical alignment », In *Terminology* 16(1), pp. 77-106.
- Marshman, E. (2008) « Expressions of uncertainty in candidate knowledge-rich contexts: A comparison in English and French specialized texts. » In *Terminology* 14(1), pp. 124-151.
- Mel'čuk, I. (2008) « Phraséologie dans la langue et dans le dictionnaire. » In Campà, À. & L. Baqué (eds), *Repères & Applications (VI)*, 2008 [un CD], Barcelona: Institut de ciències de l'educació, Universitat autònoma de Barcelona, 187-200.
- Mel'čuk, I. et Milićević, J. (à paraître) « Budalo jedna!-Type Construction in Contemporary Serbian. » In *Journal of Slavic Linguistics*.
- Mel'čuk, I. et A. Polguère. (2008) « Prédicats et quasi-prédicats sémantiques dans une perspective lexicographique ». In *LIDIL, numéro « Syntaxe et sémantique des prédicats »*, dirigé par Z. Guenchéva & Eva Novakova, n° 37, 99-114.
- Mel'čuk, I. et L. Wanner (2008) Morphological Mismatches in Machine Translation. *Machine Translation*, 22: 101-152.
- Milićević, J. (2009) « Schéma de régime: le pont entre le lexique et la grammaire. » In Blanco, X. & Pierre-André Buvet, eds. *La représentation des structures prédicat-argument. Langages* 176 (4/2009): 94-116.
- Pastor, V. et A. Alcina (à paraître) « Search Techniques in Electronic Dictionaries: a Classification for Translators », In *International Journal of Lexicography*, 23:3.
- Polguère, A. (2009) « Lexical systems: graph models of natural language lexicons. » In *Language Resources and Evaluation*, 43(1), pp. 41-55.
- Valero, E. et A. Alcina (à paraître) « Las características conceptuales en contextos ricos en conocimiento: Análisis mediante un programa de análisis cualitativo », In *Revista de Lingüística y Lenguas Aplicadas*, 5.

### 6.3 Articles dans des actes de congrès

- Barque, L. et A. Polguère (2009) « Structuration et balisage sémantique des définitions du Trésor de la langue française informatisé (TLFI) ». In *Proceedings of the fourth International Conference on Meaning-Text Theory (MTT'09)*, Montréal, 16-18 juin 2009, pp. 35-45
- Condamines, A., N. Dehaut et A. Picton (à paraître) « Utiliser des indices linguistiques pour repérer la néologie sémantique en corpus spécialisés », In *Ières Journées d'étude « Néologie sémantique et corpus : une rencontre de méthodes »*, Université de Tübingen, Allemagne, 29-30 avril 2010.
- Drouin, P. et S. R. Triantafyllos (à paraître) « Automatic extraction of transdisciplinary scientific collocations in General Scientific Language. » In *Proceedings of Language for Special Purposes 2007*.
- Estellés, A. et A. Alcina (2009), « A model of 'characteristics' formalization using Protégé-OWL », In *8<sup>e</sup> Conférence Internationale sur la Terminologie et l'Intelligence Artificielle (TIA'09)*, Toulouse, Francia, 18 al 20 de novembre de 2009.
- Hadouche, F., M.C. L'Homme, G. Lapalme et A. Le Serrec (2009) « Intégration d'informations syntaxico-sémantiques dans les bases de données terminologiques : méthodologie d'annotation et perspectives d'automatisation », In *International Workshop in Terminology and Lexical Semantics (TLS'09)*, Université de Montréal, Montréal.
- L'Homme, M.C., P. Leroyer et B. Robichaud (2010, à paraître) « Advanced Encoding for Multilingual Access in a Terminological Database – A Matter of Balance », In *Terminology and Knowledge Engineering (TKE 2010)*, 12-13 August, Dublin.
- Marengo, S. (2009) « Les dépendants syntaxiques de l'adjectif en français : vers un inventaire des relations syntaxiques de surface ». In *Proceedings of the Fourth International Conference on Meaning-Text Theory (MTT'09)*, Montréal, 16-18 juin 2009, pp. 263-273.
- Maroto, N. et A. Alcina (2009), « Conceptual relations established by processes: the case of *cocción* in a corpus of industrial ceramics », In *8<sup>e</sup> Conférence Internationale sur la Terminologie et l'Intelligence Artificielle (TIA'09)*, Toulouse, Francia, 18 al 20 de novembre de 2009.
- Marshman, E. (à paraître) « Using lexical knowledge patterns for terminology work in English and French: Some important differences. » In *Proceedings of the 16th European Symposium on Language for Special Purposes*, Hamburg, Germany, 27–31 August 2007.

- Marshman, E. et P. Van Bolderen. « Towards an integrated analysis of aligned texts: The CREATerminal approach. » In *Proceedings of Terminology and Lexical Semantics 2009*, Montreal, Canada, June 2009. CD-ROM.
- Marshman, E., M.C. L'Homme et V. Surtees (2009) « Marqueurs de la relation cause-effet : stabilité et variation dans des corpus de nature différente », In *Terminologie et intelligence artificielle* (TIA 2009). Actes. Toulouse, France.
- Marshman, E. et S. Vandaele (à paraître) Metaphorical conceptualization of associations in medical texts: An analysis in English and French. In *Proceedings of the 16th European Symposium on Language for Special Purposes*, Hamburg, Germany, 27–31 August 2007.
- Milićević, J. (2009) « C'est la définition de quel mot? Tester la validité des définitions lexicographiques pour un dictionnaire d'apprentissage. » In: Beck, D., K, gerdes, J. Milićević & A. Polguère (eds.) *Proceedings of the Fourth International Conference on Meaning-Text Theory* (MTT'09). Université de Montréal, 16-18 juin 2009; 275-285.
- Milićević, J. (2009) « Serbian Auxiliary Verbs—Syntactic Heads or Dependents? » In W. Cichocki, (ed.) *Proceedings of the 31st Annual Conference of the Atlantic Provinces Linguistics Association PAMAPLA 31*. Fredericton, 2-3 novembre 2007; pp. 43-53.
- Milićević, J. (à paraître) « Extraction de paraphrases sémantiques et lexico-syntaxiques de corpus parallèles bilingues. » In *Actes du 17 Congrès de Traitement automatique de langues naturelles*. Montréal, 19-22 July 2010.
- Milićević, J. et A. Polguère (à paraître) « Ambivalence sémantique des noms de communication langagière en français ». In *Actes du Deuxième congrès mondial de linguistique française*. New Orleans, 12-15 July 2010.
- Picton, A. (2009) « Identifier des traces d'innovation : Proposition d'une approche outillée en corpus spécialisés », In *Actes de l'atelier « Du Thème au Terme », Conférence internationale Terminologie et Intelligence Artificielle* (TIA), Toulouse, France, 21 novembre 2009.
- Pimentel, J. et M.C. L'Homme (2009) « Annotation syntaxico-sémantique de contextes spécialisés : application à la terminographie bilingue », In *8e Journées scientifiques du Réseau Lexicologie, Terminologie, Traduction « Passeurs de mots, passeurs d'espoir : lexicologie, terminologie et traduction face au défi de la diversité »*, Université nouvelle de Lisbonne, Lisbonne.
- Polguère, A. (2010) « Approche lexicaliste de la terminologie grammaticale. » In *Actes du colloque international « Enseigner la grammaire en francophonie »*, Toulouse, 16-18 février 2010.

Thibault, A. et P. Drouin (à paraître) « Le lexique de Joseph Zobel, auteur antillais : extraction semi-automatique des particularismes lexicaux ». In *Actes du XXVe Congrès International de linguistique et de philologie romane (CILPR 2007)*.

Valero, E., V. Pastor et A. Alcina (2009) « The equivalence of terms from the Internet in Spanish and French » en A. Alcina y M.-C. L'Homme (eds.) *International Workshop on Terminology and Lexical Semantics Proceedings*, Universidad de Montreal, p. 77-87. ISBN 978-2-9811149-1-4. Celebrado en la Universidad de Montreal el 19 de junio de 2009.

#### **6.4 Communications non publiées**

Ancil, D. (2009) « La norme lexicale dans une perspective didactique : positions méthodologiques pour l'analyse de problèmes lexicaux d'élèves du secondaire ». In *La norme et ses infractions en linguistique et en littérature*, du 8 au 9 mai 2009. Université Western Ontario, London.

Ancil, D. (2009) « Lexical errors in writing: a three-level analysis proposition ». In *7th IAIMTE International Conference*, du 23 au 26 juin 2009. University of Toronto.

Ancil, D. (2010) « Lexical errors in L1 writing: What are the lexical problems of high school students? ». *American Association for Applied Linguistics Annual Conference (AAAL)*, du 6 au 9 mars 2010, Atlanta, Georgie, États-Unis.

Bowker, L. et E. Marshman (2009) « Getting a better handle on translation tools before they get the better of us: Evaluating the CERTT approach. » In *IATIS Conference Panel 13: Contexts in translation education*, 8-10 juillet 2009. Melbourne, Australie.

Bowker, L. et E. Marshman (2009) « The Collection of Electronic Resources in Translation Technologies » In *Translation World*, 11-13 mai 2009. Toronto, Ontario.

Carreño, S.I. (2010) « La variation sous une perspective multilingue: analyse de variantes en corpus comparable ». Conférencière invitée. *Séminaire RALI-OLST MITACS*. Université de Montréal, 3 février 2010.

Da Sylva, L. (2010) « Extraire un “vocabulaire savant de base” français pour soutenir l'indexation automatique ». Conférencière invitée. *Séminaire RALI-OLST*, Département de linguistique et de traduction, Université de Montréal, 17 mars 2010.

Da Sylva, L. (2010) « Improving automatic indexing by actually studying how humans do it », Conférencière invitée. *McGill School of Information Science Speaker Series*, Université McGill, Montréal, 12 février 2010.

- Drouin, P. (2008) « Terminologie et technologie: état de la question, défis, évaluation, intervenants et (f)utilités », Conférencier invité. *Terminology in Advanced Management Applications* (TAMA 2008), Gatineau, Canada, 9 octobre 2008, conférencier principal.
- Drouin, P. (2009) « À la recherche des termes ou la quête initiatique du Graal », Conférencier invité. *Conférences du Laboratoire CLLE-ERSS*, Université Toulouse-Le Mirail, Toulouse, France, 20 octobre 2009.
- Drouin, P. (2009) « Approches linguistiques et statistiques à la représentation du potentiel terminologique : quelques expérimentations », Conférencier invité. *Conférences du laboratoire de Communication Langagière et Interaction Personne-Système* (CLIPS) et Groupe d'étude sur la traduction automatique (GETA). Grenoble, France, 17 décembre 2009, conférencier invité.
- Drouin, P. (2009) « Conception et utilisation du logiciel TermoStat ». Conférencier invité. *Conférences au Centre de recherche en terminologie et traduction* (CRTT), Université Lumière Lyon 2, Lyon, 7 octobre 2009, conférencier invité.
- Drouin, P. (2010) « Collocations extraction ». Conférencier invité. Université Jaume 1, Castellon, Espagne, 17 mars 2010, conférencier invité.
- Drouin, P. (2010) « The endless quest for termhood: Exploring statistical and linguistic techniques ». Conférencier invité. Universidad Politécnica de Valencia, Valencia, Espagne, 10 mars 2010, conférencier invité.
- Drouin, P. (2010) « Texto4Science », In *IXe Symposium du Conseil fédéral de terminologie*, Gatineau, Canada, 4 mai 2010, conférencier principal.
- Laroche, A. (2010) « Identification automatique de traductions de termes à partir de corpus comparables », In *Séminaires RALI-OLST*, Université de Montréal.
- Laneville, M.-E. et A. Le Serrec (2009) « DicoEnviro : Le petit frère du DiCoInfo. ». In *Séminaire RALI-OLST*. Université de Montréal, 16 septembre 2009.
- Marshman, E. (2010) « Intelligences: Terminology and Technology in Knowledge Management and Use ». Conférencière invitée. *Série Découvertes*, Faculté des Arts, Université d'Ottawa. Ottawa, Canada. 23 mars 2010.
- Marshman, E. (2010) « Happy Medium: Terminology and Terminography in the Electronic Age. » Conférence inaugurale. In *Colloque 2010 du Comité mixte sur la terminologie au Canada*. Ottawa, Canada, 18-19 février 2010.
- Marshman, E. et L. Bowker. (2009) « Attitudes of key stakeholders in translator training: Can CERTT help to turn conflict into collaboration? » In *IATIS Conference Panel 14: Translation technology and conflict*, 8-10 juillet 2009. Melbourne, Australie.

- Marshman, E. et L. Bowker. (2009) « “Before” and “after” training with translation technologies: Perceptions and needs as viewed through the lens of the CERTT project. » In *Congrès de l'Association canadienne de traductologie*, 23-25 mai 2009. Ottawa, Ontario.
- Marshman, E. et L. Bowker. (2009) « La CERTT à un CERTTain âge : Une nouvelle phase dans la vie de la Collection électronique des ressources en technologies de la traduction. » In *Actes du 78 Congrès de l'Association canadienne française pour le savoir (ACFAS)*, 11-15 mai 2009. Ottawa, Ontario.
- Mel'čuk, I. et Milićević, J. (2009) « Budalo jedna!-Type Construction in Contemporary Serbian. » In *3rd Annual Meeting of the Slavic Linguistic Society*. Zadar (Croatia), University of Zadar, 2-5 September 2009.
- Picton, A. et N. Dehaut (2010) « Variation dans les textes spécialisés et évolution des connaissances : deux exemples croisés d'analyses linguistiques dans le domaine spatial », In *78ème Congrès de l'ACFAS (Association francophone pour le savoir)*, Montréal, Canada, 10-14 mai 2010.
- Pimentel, J. et M.C. L'Homme (2009) « Annotation syntaxico-sémantique de contextes spécialisés : application à la terminographie bilingue. » In *Séminaires OLST-RALI*, Université de Montréal, 11 novembre 2009.
- Tremblay, O. (2010). « Quels savoirs pour la didactique du lexique ? » In *Séminaires Mididactiques*. Université de Montréal, 11 février 2010.
- Tremblay, O. et M. Mottet (2009). « L'entretien sociolinguistique de groupe pour faire émerger les représentations sur la langue ». In *Méthodes d'enseignement-évaluation et méthodologie de recherche en oral et didactique de l'oral au 77e congrès de l'Association francophone pour le savoir (ACFAS)*, Ottawa, Canada.
- Tremblay, O. et M. Mottet (2010). « Pratiques effectives du français québécois oral standard chez de futurs maîtres au primaire en contexte de stage. » In *Discours oraux des élèves et des enseignants : représentations, approches didactiques, pratiques rapportées et pratiques effectives. 78e congrès de l'Association francophone pour le savoir (ACFAS)*, Montréal, Canada.
- Tremblay, O. et M. Mottet (2010). « Représentations du français oral québécois et du rôle professionnel chez de futurs enseignants du primaire. » In *Discours oraux des élèves et des enseignants : représentations, approches didactiques, pratiques rapportées et pratiques effectives. 78e congrès de l'Association francophone pour le savoir (ACFAS)*, Montréal, Canada.

## 6.5 Autres

L'Homme, M.C. (dir.) (2007) *DicoEnviro. Dictionnaire fondamental de l'environnement*.  
Version Web : <http://olst.ling.umontreal.ca/dicoenviro>

L'Homme, M.C. (dir.) (2007) *DicoInfo. Dictionnaire fondamental de l'informatique et de l'Internet*. Version Web : <http://olst.ling.umontreal.ca/dicoinfo>

Barque L. et A. Polguère (2009) *Guide des annotateurs pour la structuration des définitions du TLFi*. Rapport technique, Département de linguistique et de traduction, Université de Montréal, 16 pages.  
Version Web : <http://olst.ling.umontreal.ca/pdf/guideAnnoTLFi.pdf>

Pimentel, J. (à paraître, 2010) Compte rendu de l'ouvrage *Topics in Language Resources for Translation and Localisation* de E. Rodrigo (ed.) In META.

Robichaud, B. (2010) *OLSTER*. Recherches dans les bases de données lexicales électroniques de l'OLST. Site Web : <http://olst.ling.umontreal.ca/olster>