

Bourse de thèse en lexicologie/lexicographie informatisée

Dans le cadre des contrats doctoraux (allocations de recherches) portant sur des thématiques prioritaires, le Laboratoire ATILF (Analyse et Traitement Informatique de la Langue Française) du CNRS, UMR 7118, et l'École Doctorale « Langages, Temps, Sociétés » de l'Université Nancy 2 proposent une bourse de recherche de thèse dans le domaine de la lexicologie/lexicographie informatisée.

Date de début Septembre 2010

Lieu Laboratoire ATILF (Nancy)

Durée 3 ans

Intitulé du sujet de thèse proposé

Graphes lexicaux tissés par les fonctions lexicales Sens-Texte

Directeur de thèse

Alain Polguère

alain.polguere@univ-nancy2.fr

Descriptif de la recherche

On entend par *graphe lexical* un type particulier de modèle formel du lexique des langues. Dans un tel modèle, la connaissance lexicale est représentée sous forme d'un graphe où les lexies — nœuds du graphe — sont connectées entre elles par un ensemble potentiellement très riche de liens significatifs — arcs du graphe. Toute lexie L sémantiquement pleine peut contrôler deux types majeurs de liens lexicaux.

1. Les liens paradigmatiques unissent L à d'autres lexies qui entretiennent avec elle une relation sémantique forte. Il s'agit, par exemple, de synonymes, antonymes, conversifs, nominalisations, verbalisations, etc.
2. Les liens syntagmatiques unissent L à d'autres lexies (ou à des syntagmes) avec lesquelles elle forme des collocations — c'est-à-dire, des expressions phraséologisées, mais compositionnelles, du type *pousser un cri*, *poser une question*, *sommeil lourd*... Il s'agit donc de liens vers des verbes supports, des intensificateurs, etc.

La Lexicologie Explicative et Combinatoire ou LEC (Mel'čuk et al., 1995), la branche lexicale de la théorie Sens-Texte, a développé un outil formel servant à modéliser les liens lexicaux paradigmatiques et syntagmatiques qui se manifestent de façon récurrente dans les langues : il s'agit du système des fonctions lexicales standard. Ce dernier a été utilisé de façon méthodique en lexicologie théorique et descriptive (Mel'čuk et al., 1984, 1988, 1992, 1999; Fontenelle, 1997; Polguère, 2009).

Le sujet de thèse proposé vise à examiner la structure relationnelle des lexiques lorsque celle-ci est modélisée sur la base du système des fonctions lexicales. Il s'agira, d'une part, de trouver une représentation optimale formelle d'un graphe lexical tissé par les fonctions lexicales Sens-Texte et, d'autre part, d'en examiner les propriétés formelles et d'exploiter l'éclairage qu'il peut apporter sur l'organisation du lexique des langues.

La recherche, bien qu'à visée avant tout théorique, inclura une expérimentation menée sur le lexique du français, dans la continuité des travaux déjà réalisés dans le cadre de la LEC.

Enjeux scientifiques

La problématique de la construction de lexiques formels occupe une place de plus en plus importante en linguistique et en traitement automatique des langues (TAL). On remarque que les structures de données utilisées pour construire et étudier de tels lexiques s'éloignent de plus en plus de la structure « textuelle » qui est celle des dictionnaires — y compris les dictionnaires électroniques —, pour se rapprocher de véritables structures de graphes — voir, par exemple, WordNet (Fellbaum, 1998), FrameNet (Baker et al., 2003) ou les Petits Mondes Lexicaux (Gaume, 2004).

S'il existe des raisons pratiques pour s'orienter vers des modèles lexicaux du type graphes, raisons liées notamment à la plus grande calculabilité de ces modèles, il existe aussi des raisons théoriques profondes, liées à la nature relationnelle de la connaissance lexicale (Aitchison, 2003).

Profil du candidat

Le candidat devra avoir un solide bagage en linguistique formelle et manifester de l'intérêt pour le problème de l'interaction entre les deux composantes structurales majeures de la langue : le lexique et la grammaire (sémantique, syntaxe de dépendance, morphologie, phonologie).

Une formation additionnelle en informatique ou en théorie des graphes est un atout certain. La langue de travail peut être le français ou l'anglais.

Modalités de soumission d'une candidature

1. Il faut envoyer un CV au directeur de thèse **le plus tôt possible** (avant le 18 juin 2010). Ce dernier aidera le candidat à préparer un prédossier, qui doit être soumis au Conseil de direction scientifique (CDS) de l'ATILF **avant le 21 juin 2010**.
2. Le dossier complet de candidature devra parvenir à l'École Doctorale « Langages, Temps, Sociétés » de Nancy 2 **le vendredi 25 juin 2010 à 12h** au plus tard. Ce dossier doit comporter un CV détaillé, les relevés de notes (licence et master) et tout autre document jugé pertinent.
3. Les candidats à un contrat doctoral seront auditionnés par l'École Doctorale le lundi 5 juillet 2010. Cette audition sera décisionnelle. Une audition de préparation par le CDS de l'ATILF est prévue une semaine avant, le lundi 28 juin 2010.

Pour plus d'informations sur cette offre de bourse, prendre directement contact avec le directeur de thèse (alain.polguere@univ-nancy2.fr).

Références

- Jean Aitchison. *Words in the Mind: An Introduction to the Mental Lexicon*. Blackwell, Oxford, UK, 3^e édition, 2003.
- Collin F. Baker, Charles J. Fillmore et Beau Cronin. The Structure of the FrameNet Database. *International Journal of Lexicography*, 16(3):281–296, 2003.
- Christiane Fellbaum, dir. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, MA, 1998.
- Thierry Fontenelle. *Turning a bilingual dictionary into a lexical-semantic database*. Niemeyer, Tübingen, 1997.
- Bruno Gaume. Balades aléatoires dans les Petits Mondes Lexicaux. *I3 Information Interaction Intelligence*, 4(2):39–96, 2004.
- Igor Mel'čuk, André Clas et Alain Polguère. *Introduction à la lexicologie explicative et combinatoire*. Duculot, Paris/Louvain-la-Neuve, 1995.
- Igor Mel'čuk et al. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques. Volumes I–IV*. Les Presses de l'Université de Montréal, Montréal, 1984, 1988, 1992, 1999.
- Alain Polguère. Lexical systems: graph models of natural language lexicons. *Language Resources and Evaluation*, 43(1):41–55, mars 2009.