

Intitulé du sujet en français :	<i>Application du modèle lexicologique des grands réseaux lexicaux à la structure multidimensionnelle des terminologies scientifiques : le cas du lexique de la chimie</i>
Intitulé du sujet en anglais :	<i>Applying the lexicological model of large lexical networks to the multidimensional structure of scientific terminologies: the case of the lexicon of chemistry</i>
Domaines scientifiques de la thèse :	lexicologie théorique et descriptive, terminologie de la chimie, grands réseaux lexicaux

Directeur de thèse :	Alain POLGUÈRE
Courriel :	alain.polguere@univ-lorraine.fr
Codirectrice de thèse :	Francesca INGROSSO
Courriel :	francesca.ingrosso@univ-lorraine.fr

Argumentaire scientifique (descriptif du sujet, enjeux scientifiques) :

### **Descriptif du sujet<sup>1</sup>**

Ce projet de recherche doctorale interdisciplinaire est motivé par un double constat : d'une part, les approches traditionnelles d'étude et de représentation des systèmes terminologiques scientifiques reposent essentiellement sur le recours à des modèles taxinomiques – cf. les ontologies informatiques fondées sur des hiérarchies de classes conceptuelles (Hirst 2009) ; d'autre part, les études contemporaines sur l'organisation du lexique de langue générale (lexicologie, psycholinguistique, etc.) tendent à s'accorder sur un mode de structuration en « réseaux lexicaux » non hiérarchiques et non taxinomiques (Motter et coll. 2002 ; Aitchison 2012 ; Polguère 2016). La recherche proposée repose sur l'hypothèse que les terminologies scientifiques, puisqu'elles fonctionnent dans les textes en interaction avec le lexique de langue générale, doivent posséder une structure homomorphe avec celle du lexique général, avec lequel elles fusionnent au sein de la langue (L'Homme & Polguère 2008).

La recherche doctorale vise à théoriser la notion de structure non taxinomique des terminologies en utilisant le modèle des Systèmes Lexicaux (Polguère 2014) – graphes lexicaux dits *petits-mondes* (Gaume 2008) développés au laboratoire ATILF –, dans le cadre d'une étude descriptive en profondeur de la terminologie de base de la chimie, pour les langues anglaise et française (McNaught & Wilkinson 1997 ; Peraldi 2011 ; Zanola 2014). Une attention particulière sera portée à l'interface entre langue scientifique et langue générale, autour des unités lexicales générales et des termes de la chimie qui établissent cette jonction (*réaction, synthèse, gaz...*). La méthodologie sera ancrée dans la modélisation lexicographique des termes (sens, caractéristiques grammaticales, connexions lexicales sémantiques et phraséologie), avec un recours systématique aux corpus textuels de chimie et à l'interaction avec les scientifiques du domaine. L'approche théorique et descriptive sous-jacente sera celle de la Lexicologie Explicative et Combinatoire (Mel'čuk et coll. 1995), qui a été amplement appliquée à l'étude de vocabulaires terminologiques, tels que ceux de l'informatique et de l'environnement (L'Homme 2012).

<sup>1</sup> Voir *Bibliographie indicative* ci-dessous, pour les références bibliographiques.

Un travail préparatoire sur la terminologie de la chimie a été effectué par A. Polguère dans le cadre du projet PEPS Mirabelle STRÉCH de F. Ingrosso (SRSMC) – codirectrice de la thèse –, en collaboration avec deux autres membres du laboratoire de chimie SRSMC. Ce projet a mené notamment à la construction de deux corpus anglais et français, dont la taille est suffisamment significative (3 millions de mots-occurrences pour chacun des corpus) pour permettre une première estimation des nomenclatures terminologiques à considérer et pour effectuer des expérimentations sur la description terminographique du vocabulaire de la chimie – pour une présentation de l’approche adoptée et des travaux réalisés, voir Ingrosso & Polguère (2015). La recherche doctorale inclura notamment un travail d’extension et d’affinage de ces corpus ainsi que leur traitement au moyen d’outils d’extraction de termes et d’analyse terminologique (Drouin 2009).

### **Enjeux scientifiques et pertinence du sujet pour le laboratoire ATILF**

Sur le plan théorique, la recherche vise à apporter une solution au problème de la modélisation de la multidimensionnalité inhérente à l’organisation des terminologies (Bowker 1997), tout en rendant compte de l’interdépendance entre lexique de langue générale et lexique terminologique. Sur le plan pratique, la thèse débouchera sur des modèles terminologiques en français et en anglais pour la chimie, modèles qui seront des outils destinés aux scientifiques aussi bien qu’aux enseignants. Les chimistes sont aujourd’hui fréquemment amenés à faire face à des thématiques d’intérêt public et, donc, à communiquer dans des contextes où le matériau linguistique/notionnel qui structure leur discipline est employé et interprété en dehors de la seule activité scientifique. Ils sont ainsi de plus en plus conscients de l’importance d’une approche non isolationniste de la terminologie de leur domaine et de la nécessité d’une amélioration des stratégies d’enseignement de celle-ci (Kermen 2016). La recherche proposée est donc destinée à avoir une résonance non seulement dans le domaine de la recherche en lexicologie et terminologie, mais aussi auprès de la communauté des chimistes.

La thèse se situe dans la thématique des études lexicales, qui sont au cœur du projet scientifique du laboratoire ATILF. Plus particulièrement, la recherche sera ancrée dans les travaux menés sur les grands réseaux lexicaux par l’axe « Lexicologie théorique et descriptive » de l’équipe Lexique, axe qui est sous la responsabilité du directeur de la thèse – A. Polguère. Une exploitation intensive sera faite des modèles lexicaux développés dans le cadre du travail sur les Réseaux Lexicaux du Français et de l’Anglais. En retour, la recherche doctorale alimentera ces ressources en données sur les terminologies anglaises et françaises de la chimie.

La recherche sur les grands réseaux lexicaux intéresse un nombre significatif de membres de l’ATILF, notamment dans le cadre des recherches menées sur les corpus terminologiques. Il s’agit aussi d’un domaine de recherche à fort potentiel applicatif (traitement automatique de la langue, enseignement...).

Finalement, cette thèse est tout particulièrement intéressante pour le laboratoire du fait de son aspect fortement interdisciplinaire : lexicologie/terminologie, chimie et traitement automatique de la langue (bases de données lexicales et leurs applications).

### **Présentation des encadrants**

Le directeur de thèse – A. Polguère – est un spécialiste de la lexicologie formelle informatisée et de ses applications, tant en traitement automatique de la langue qu’en enseignement et apprentissage du lexique.

La codirectrice – F. Ingrosso – mène ses recherches en chimie théorique dans les domaines suivants : modélisation, dynamique moléculaire ; chimie quantique ; solvatation en phase liquide et dans les solvants supercritiques. Elle soutiendra son HDR d’ici l’été 2017.

Même si les deux encadrants appartiennent à des champs disciplinaires très distincts, la thèse proposée est une suite logique d’une collaboration interdisciplinaire déjà bien amorcée, notamment à travers le projet préparatoire STRÉCH.

Profil attendu du/de la candidat.e :

Le/la candidat.e devra posséder un master en sciences du langage, avec une spécialisation dans l'étude du lexique : lexicologie, lexicographie, terminologie... Il/elle devra avoir un intérêt poussé pour la formalisation des données lexicales et l'utilisation des outils informatiques. Une expérience de l'exploitation des corpus linguistiques est souhaitable.

Compte tenu de la nature des données linguistiques étudiées, le/la candidat.e devra posséder une connaissance de la problématique de l'analyse terminologique (langage scientifique, technique, etc.). Il faudra de plus qu'il/elle maîtrise le français et l'anglais – langue maternelle pour au moins une de ces deux langues.

Bibliographie indicative :

- Aitchison, J. (2012). *Words in the Mind: An Introduction to the Mental Lexicon*. 4<sup>th</sup> edition, Oxford : Wiley-Blackwell.
- Bowker, L. (1997). Multidimensional Classification of Concepts and Terms. Dans S. E. Wright and G. Budin (dir.), *Handbook of Terminology Management: Volume 1: Basic Aspects of Terminology Management* (pp. 133–143). Amsterdam/Philadelphia : Benjamins.
- Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology*, **9(1)**, 99–117.
- Gaume, B. (2008). Mapping the Forms of Meaning in Small Worlds. *Journal of Intelligent Systems*, **23**, 848–862.
- Hirst, G. (2009). Ontology and the Lexicon. Dans S. Staab & R. Studer (dir.), *Handbook on Ontologies* (pp. 269–292). 2<sup>nd</sup> ed., Berlin Heidelberg : Springer.
- Ingrosso, F., Polguère, A. (2015). How Terms Meet in Small-World Lexical Networks: The Case of Chemistry Terminology. Dans T. Poibeau & P. Faber (dir.), *Proceedings of the 11<sup>th</sup> International Conference on Terminology and Artificial Intelligence (TIA 2015)* (pp. 167–171). Granada.
- Kermen, I. (2016). Usage et apprentissage des mots en chimie : difficultés, ambiguïtés, obstacles. *L'actualité chimique*, **407**, 34–36.
- L'Homme, M.-C. (2012). Using ECL (Explanatory Combinatorial Lexicology) to discover the lexical structure of specialized subject fields. Dans Apresjan, J. et coll. (dir.), *Words, Meanings and other Interesting Things. A Festschrift in Honour of the 80<sup>th</sup> Anniversary of Professor Igor Alexandrovič Mel'čuk* (pp. 378–390). Moscow : Jazyki slavjanskoj kultury Publishers.
- L'Homme, M.-C., Polguère, A. (2008). Mettre en bons termes les dictionnaires spécialisés et les dictionnaires de langue générale. Dans F. Maniez & P. Dury (dir.), *Lexicologie et terminologie : histoire de mots. Hommage à Henri Béjoint* (pp. 191–206). Lyon : Travaux du CRTT.
- McNaught, A. D., Wilkinson, A. (1997) *IUPAC. Compendium of Chemical Terminology. 2nd ed. (the "Gold Book")*, Oxford: Blackwell Scientific Publications. [XML on-line corrected version: <http://goldbook.iupac.org> (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. ISBN 0-9678550-9-8. doi:10.1351/goldbook]
- Mel'čuk, I., Clas, A., Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Paris/Louvain-la-Neuve : Duculot.
- Motter, A. E., de Moura, A. P.S., Lai, Y.-C., Dasgupta, P. (2002). Topology of the conceptual network of language. *Physical Review E*, **65**, 065102-1–4.
- Peraldi, S. (2011). *Indétermination terminologique et multidimensionnalité dans le domaine de la chimie organique*. Thèse de doctorat, Université Paris Diderot (Paris 7), Paris.
- Polguère, A. (2014). From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography*, **27**, 396–418.
- Polguère, A. (2016). La question de la géométrie du lexique. *SHS Web of Conferences*, **27**, 01002, Actes du « 5<sup>e</sup> Congrès Mondial de Linguistique Française (CMLF 2016) ».
- Zanola, M. T. (2014). *La nomenclature chimique. Arts et métiers au XVIII<sup>e</sup> siècle. Études de terminologie diachronique* (pp. 113–128). Paris : L'Harmattan.