

Intitulé du sujet en français	<i>Étude multilingue du lexique de la chimie à l'interface entre terminologie et langue générale</i>
Intitulé du sujet en anglais	<i>Multilingual study of the lexicon of chemistry at the interface of terminology and general language</i>
Mots clés	lexicologie contrastive, lexicographie informatisée, terminologie de la chimie, grands réseaux lexicaux

Directeur de thèse **Alain POLGUÈRE** (Université de Lorraine, CNRS, ATILF)
 alain.polguere@univ-lorraine.fr
<https://apps.atilf.fr/homepages/polguere/fr/accueil/>

Codirectrice de thèse **Francesca INGROSSO** (Université de Lorraine, CNRS, LPCT)
 francesca.ingrosso@univ-lorraine.fr
<https://cpc2016nancy.wixsite.com/fingrosso>

Contexte

Ce projet de recherche s'appuie sur un double constat : d'une part, les approches traditionnelles de représentation des *terminologies 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 (Aitchison, 2012; Polguère, 2016). La recherche proposée fait 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 et Polguère, 2008).

Les modèles terminologiques, dans leur immense majorité, traitent les terminologies de façon isolationniste et ne sont pas en mesure de rendre compte des types de connexions qu'elles entretiennent avec la langue générale et des conséquences pratiques de ces connexions sur la dynamique lexicale (évolution des terminologies), ainsi que sur l'enseignement et l'apprentissage des terminologies. La recherche doctorale étudiera ce problème dans une perspective à la fois théorique (lexicologique) et descriptive (lexicographique/terminographique) en se consacrant au cas de la *terminologie de la chimie*.

La réflexion sur le langage de la chimie a une longue et riche histoire (Zanola, 2014, pp. 113–128). Pourtant, l'essentiel des efforts de normalisation et rationalisation de cette terminologie s'est concentré sur le nommage des composés chimiques, puisque la chimie, de par la nature même de la discipline, engendre de façon continue de nouvelles molécules et, donc, de nouveaux noms pour les désigner. L'importance d'une utilisation correcte de la terminologie en chimie a mené, à partir de 1919, à la création d'un organisme international – *IUPAC : International Union for Pure and Applied Chemistry*¹ –, dont une des activités consiste à développer des règles pour la nomenclature des molécules, afin d'éviter des ambiguïtés dans les définitions et harmoniser les nouvelles propositions lors de la découverte d'une molécule.

Malgré l'intérêt évident que les chimistes portent aux questions d'utilisation des termes dans la pratique de leur discipline, les exemples d'études systématiques du lexique de la chimie restent étonnamment rares. Citons les travaux de thèse S. Peraldi (Peraldi, 2011), dans lesquels elle analyse le langage de la chimie organique, en mettant en évidence son caractère indéterminé et multidimensionnel. Par l'élaboration d'un corpus spécialisé (en anglais), Peraldi propose des pistes pour la construction d'un modèle terminologique qui pourrait assister les chimistes organiciens dans leurs efforts de communication scientifique. En effet, dans la rédaction d'articles comme dans l'expression orale, tout(e) scientifique est confronté(e)

1. <https://iupac.org>

aux difficultés liées à la communication de concepts qui sont souvent très techniques et spécifiques d'un sous-domaine (par ex., chimie organique, chimie des matériaux, chimie physique). D'autres difficultés ressortent dans un contexte pédagogique : nécessité d'utiliser une expression claire, d'éviter des définitions ambiguës, de privilégier l'illustration de concepts simples par des termes bien choisis.

Objectifs et pertinence pour le laboratoire

Sur le plan théorique, la recherche vise, d'une part, à apporter une solution au problème de la modélisation formelle et rigoureuse de la *multidimensionnalité* inhérente à l'organisation des terminologies (Bowker, 1997), c'est-à-dire le fait que les termes peuvent être appréhendés et les terminologies parcourues selon de multiples axes. D'autre part, et de façon liée, la recherche vise à rendre compte de l'*interdépendance entre lexicologie de langue générale et lexicologie terminologique*. Une des problématiques centrales de la description des terminologies étant celle de leur traduction, la recherche adoptera une *perspective multilingue*, en s'appuyant sur les travaux de lexicologie théorique et descriptive menés à l'ATILF sur le français, l'anglais et le russe.

Sur le plan pratique, la thèse débouchera sur des *modèles terminologiques multilingues* de la chimie pour les trois langues typologiquement distinctes considérées. Ces modèles seront des outils destinés aux scientifiques aussi bien qu'aux enseignants en chimie. 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 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. La recherche proposée est de ce fait 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.

En regard de tous les challenges posés par la question terminologique dans le domaine de la chimie, il faut noter que les efforts de l'IUPAC (cf. plus haut, Section « Contexte ») sont principalement orientés vers la normalisation de la terminologie très prolifique des composés chimiques. En comparaison, peu d'attention a été portée sur une meilleure normalisation de ce que l'on appellera la *terminologie fondamentale de la chimie*, c'est-à-dire la terminologie qui forme le noyau notionnel de la discipline. Cette terminologie est celle qui pose le plus de problèmes quant à l'interaction avec la langue générale, par exemple dans les contextes pédagogiques. La terminologie fondamentale de la chimie représente, de plus, le socle sur lequel s'appuie toute la terminologie dénomminative des substances. La recherche doctorale proposée va donc porter spécifiquement sur la nomenclature fondamentale de la chimie, nomenclature qu'il s'agira, dans un premier temps, de déterminer de façon rigoureuse, avant d'en mener l'étude et la description.

Le projet se situe dans la thématique des études lexicales, qui sont au cœur du projet scientifique du laboratoire ATILF. Il présente l'originalité d'aborder le sujet du *rapport entre terme et non-terme* dans le cadre des travaux menés à l'ATILF par A. Polguère sur les grands réseaux lexicaux. Une exploitation intensive sera faite des modèles lexicaux développés au laboratoire depuis plusieurs années : les réseaux lexicaux du français et de l'anglais (Gader et al., 2014) et, dans le cadre d'une collaboration avec le laboratoire CREE de l'INALCO Paris (Projet *LEC-ru*), le réseau du russe. En retour, la recherche doctorale alimentera ces ressources en données sur les terminologies anglaise, française et russe de la chimie.

Méthodologie

La recherche doctorale entreprendra de théoriser la notion de structure non taxinomique des terminologies en utilisant les modèles des *Systèmes Lexicaux* (Polguère, 2014) – graphes lexicaux dits *petits-mondes* ou *grands graphes de terrain* (Gaume, 2008) – développés au laboratoire ATILF². La réflexion théorique s'appuiera sur une étude descriptive en profondeur de la terminologie fondamentale de la chimie (cf. section précédente) pour, les langues française, anglaise et russe. Comme mentionné plus haut, l'attention se portera sur l'interface entre langue générale et langue scientifique, autour des unités lexicales générales – les non-termes – et des termes de la chimie qui établissent cette jonction (*réaction, synthèse, gaz...*).

La détermination d'une terminologie fondamentale de la chimie est, de ce fait, un préalable à toute étude descriptive et théorique. Un travail préparatoire a été effectué par les deux encadrants de la thèse

2. <https://apps.atilf.fr/homepages/polguere/fr/lexical-systems/>

dans le cadre du projet PEPS Mirabelle STRÉTCH de F. Ingrosso, en collaboration avec deux autres membres l'ancien laboratoire de chimie SRSMC, M. F. Ruiz-López (maintenant, LPCT) et C. Comoy (maintenant, L2CM). Ce projet a mené notamment à la construction de deux corpus français et anglais, 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 et 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, 2003). Le même travail de construction et exploitation de corpus sera entrepris pour la langue russe. Cette étape de la recherche mènera à la finalisation d'une première version d'une terminologie fondamentale multilingue de la chimie, qui pourra comporter plusieurs centaines de termes, dans chaque langue prise en charge. La terminologie identifiée sera alors modélisée selon l'approche des Systèmes Lexicaux.

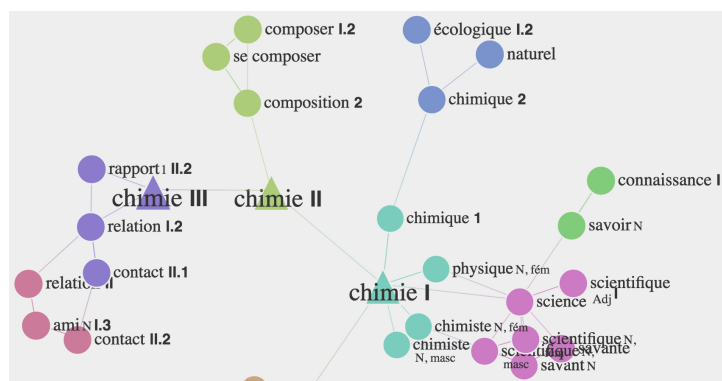


FIGURE 1 – Trois sens de CHIMIE dans le RL-fr

Ces derniers possèdent des propriétés formelles qui en font des outils de modélisation puissants pour faire émerger des aires sémantiques au sein de leur structure relationnelle. Les procédures d'*analyse proxémique* (Gaume et al., 2006) fonctionnent déjà sur les Systèmes Lexicaux que nous développons. La Figure 1 illustre ce fait, avec la visualisation d'*espaces sémantiques* automatiquement identifiés dans la polysémie du vocable CHIMIE dans le Réseau Lexical du Français (RL-fr) : sens I 'discipline scientifique' [*chercheuse en chimie théorique*]; sens II, métonymique [*chimie d'un acide*]; sens III, métaphorique [*chimie sentimentale entre deux personnes*].

Dans cette visualisation, les espaces sémantiques apparaissent comme des *clusters* de nœuds lexicaux de même couleur. La proximité sémantique relative entre nœuds/*clusters* est calculée en fonction de la topologie du graphe et est représentée dans la visualisation par leur distance géométrique. Les propriétés formelles des Systèmes Lexicaux permettront d'étudier la façon dont une terminologie donnée – dans notre cas, celle de la chimie – pénètre la langue générale et, réciproquement, est pénétrée par elle. Le projet fait l'hypothèse que la mise en réseau de façon fusionnelle des deux domaines lexicaux (de langue générale et terminologique) permettra de faire émerger des propriétés des terminologies quant à leur relation avec la langue générale. Pour cela, il conviendra de mener un travail poussé de description lexicographique des termes : sens, caractéristiques grammaticales, connexions lexicales sémantiques et phraséologiques (expressions idiomatiques impliquant les termes), avec un recours systématique aux corpus textuels de chimie et à l'interaction avec les spécialistes du domaine du laboratoire de chimie LPCT³.

L'approche théorique et descriptive sera celle de la *Lexicologie Explicative et Combinatoire* (Mel'čuk et al., 1995). Le travail lexicographique sera mené de façon comparative – pour les trois langues prises en compte – afin d'identifier les 1) les divergences de structuration terminologique entre langues et 2) les divergences quant à l'interconnexion entre terminologie et langue générale.

Le développement du projet bénéficiera d'une véritable immersion dans les questions de communication en chimie par l'intermédiaire de la codirection, impliquant le laboratoire de chimie LPCT. Dans une toute première partie de la thèse, le/la doctorant(e) sera mené(e) à se former aux bases notionnelles de la chimie. Dans la phase d'élaboration de la nomenclature fondamentale de la chimie, des discussions régulières avec des chimistes seront nécessaires. Pour élargir le spectre des compétences, des scientifiques d'autres laboratoires (L2CM pour la chimie moléculaire, LPCME pour la chimie de l'environnement, CRM2 pour la chimie du solide, IJL pour la chimie des matériaux) seront invités/consultés régulièrement. Dans ces interactions, les efforts seront concentrés, dans un premier temps, sur l'analyse des corpus par rapport à leur représentativité (choix de revues et d'ouvrages, sous-domaines de la chimie, périodes

3. <http://lpct.univ-lorraine.fr/index.html>

examinées). Ensuite, un travail de fond sera consacré à l'étude des définitions terminologiques issues du traitement des corpus développés. Finalement, un point important de cette approche interdisciplinaire est que, en fréquentant le LPCT, le/la doctorant(e) sera sollicité(e) pour participer aux séminaires, pour interagir avec les scientifiques invités, pour intervenir lors de l'organisation de manifestations scientifiques (colloques, ainsi qu'événements de médiation scientifique tels que la Fête de la Science). Cela apportera des connaissances spécifiques vis-à-vis des pratiques d'utilisation de la langue dans le domaine de la chimie et des techniques de communication utilisées. Le/la doctorant(e) pourra concrètement examiner les problèmes qui peuvent se présenter dans un contexte de communication.

Profil attendu du candidat ou de la candidate

Le/la candidat(e) devra posséder une excellente formation en linguistique et, particulièrement, en lexicologie et/ou terminologie – par exemple, il/elle pourrait être issu(e) du programme d'une des sept universités européennes partenaires dans le master Erasmus mundus *European Master in Lexicography* (EMLex)⁴.

Il/elle devra posséder une bonne connaissance des approches formelles et informatisées en sciences du langage, ainsi qu'un intérêt marqué pour les sciences exactes et expérimentales – la familiarité avec le domaine de la chimie étant un atout sans être aucunement une nécessité.

Compétences linguistiques : la maîtrise du français, de l'anglais et du russe est nécessaire.

Bibliographie

- Jean Aitchison. *Words in the Mind: An Introduction to the Mental Lexicon*. Wiley-Blackwell, Oxford, UK, 4th édition, 2012.
- Lynne Bowker. Multidimensional Classification of Concepts and Terms. Dans Sue Ellen Wright et Gerhard Budin, dir., *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*, pages 133–143. John Benjamins, Amsterdam/Philadelphia, 1997.
- Patrick Drouin. Term extraction using non-technical corpora as a point of leverage. *Terminology*, 9(1):99–117, 2003.
- Nabil Gader, Sandrine Ollinger et Alain Polguère. One Lexicon, Two Structures: So What Gives? Dans Heili Orav, Christiane D. Fellbaum et Piek Vossen, dir., *Proceedings of the Seventh Global Wordnet Conference (GWC2014)*, pages 163–171, Tartu (Estonie), 25–29 janvier 2014. Global WordNet Association.
- Bruno Gaume. Mapping the Forms of Meaning in Small Worlds. *Journal of Intelligent Systems*, 23:848–862, 2008.
- Bruno Gaume, Fabienne Venant et Bernard Victorri. Hierarchy in Lexical Organisation of Natural Languages. Dans Denise Pumain, dir., *Hierarchy in Natural and Social Sciences*, Methodos series 3, pages 121–142. Springer, Dordrecht, 2006.
- Graeme Hirst. Ontology and the Lexicon. Dans Steffen Staab et Rudi Studer, dir., *Handbook on Ontologies*, International Handbooks on Information Systems, pages 269–292. Springer, Berlin Heidelberg, 2nd édition, 2009.
- Francesca Ingrosso et Alain Polguère. How Terms Meet in Small-World Lexical Networks: The Case of Chemistry Terminology. Dans Thierry Poibeau et Pamela Faber, dir., *Proceedings of the 11th International Conference on Terminology and Artificial Intelligence (TIA 2015)*, pages 167–171, Granada, 2015.
- Marie-Claude L'Homme et Alain Polguère. Mettre en bons termes les dictionnaires spécialisés et les dictionnaires de langue générale. Dans François Maniez et Pascaline Dury, dir., *Lexicologie et terminologie : histoire de mots. Hommage à Henri Béjoint*, pages 191–206. Travaux du CRTT, Lyon, 2008.
- Igor Mel'čuk, André Clas et Alain Polguère. *Introduction à la lexicologie explicative et combinatoire*. 256 pages. Duculot, Paris/Louvain-la-Neuve, 1995.
- Sandrine Peraldi. *Indétermination terminologique et multidimensionnalité dans le domaine de la chimie organique*. Thèse de doctorat, Université Paris Diderot (Paris 7), Paris, 2011.
- Alain Polguère. From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography*, 27(4):396–418, 2014. Special Issue: Dictionaries and the Digital Revolution: A Focus on Users and Lexical Databases.
- Alain Polguère. La question de la géométrie du lexique. *SHS Web of Conferences*, 27:01002, 2016. Actes du « 5^e Congrès Mondial de Linguistique Française (CMLF 2016) ».
- Maria Teresa Zanola. *Arts et métiers au XVIII^e siècle. Études de terminologie diachronique*. L'Harmattan, Paris, 2014.

4. <https://www.emlex.phil.fau.eu>