

Principes de modélisation systémique des réseaux lexicaux *

Alain Polguère

ATILF, CNRS & Université de Lorraine, 44 av. de la Libération, BP 30687, 54063 Nancy Cedex
alain.polguere@univ-lorraine.fr

Résumé. Nous présentons une approche de la construction manuelle des ressources lexicales à large couverture fondée sur le recours à un type particulier de réseau lexical appelé *système lexical*. En nous appuyant sur l'expérience acquise dans le cadre de la construction du Réseau Lexical du Français (RL-fr), nous offrons tout d'abord une caractérisation formelle des systèmes lexicaux en tant que graphes d'unités lexicales de type « petits mondes » principalement organisés à partir du système des fonctions lexicales Sens-Texte. Nous apportons ensuite des arguments pour justifier la pertinence du modèle proposé, tant du point de vue théorique qu'applicatif.

Abstract. We introduce a new approach for manually constructing broad-coverage lexical resources based on a specific type of lexical network called *lexical system*. Drawing on experience gained from the construction of the French Lexical Network (fr-LN), we begin by formally characterizing lexical systems as “small-world” graphs of lexical units that are primarily organized around the system of Meaning-Text lexical functions. We then give arguments in favor of the proposed model that are both theory- and application-oriented.

Mots-clés : système lexical, base de données lexicale, structure du lexique, réseau lexical, graphe petit monde, proxémie, Lexicologie Explicative et Combinatoire, fonction lexicale, Réseau Lexical du Français (RL-fr).

Keywords: lexical system, lexical database, structure of the lexicon, lexical network, small-world graph, proxemy, Explanatory Combinatorial Lexicology, lexical function, French Lexical Network (fr-LN).

1 Introduction

Cet article rend compte des principes de structuration de la connaissance lexicale mis en pratique dans le cadre de la construction manuelle, par une équipe de lexicographes, du Réseau Lexical du Français, désormais RL-fr. Cette ressource est élaborée suivant le cadre théorique et descriptif général de la Lexicologie Explicative et Combinatoire (Mel'čuk *et al.*, 1995; Mel'čuk, 2006). Elle se distingue cependant radicalement des travaux lexicographiques antérieurement effectués selon cette approche par le fait qu'elle repose sur un modèle non dictionnaire et non textuel appelé *système lexical*.

La première étape de construction du RL-fr, qui doit s'achever fin 2014, s'est effectuée dans le cadre des travaux du projet RELIEF, menés au laboratoire ATILF du CNRS. Plusieurs publications traitent de l'implémentation du RL-fr, de son édition lexicographique et de sa croissance (Lux-Pogodalla & Polguère, 2011; Gader *et al.*, 2012; Polguère & Sikora, 2013). Nous nous intéressons ici spécifiquement au type particulier de modélisation systémique des lexiques qu'il implémente, en nous concentrant sur deux points : nature formelle du système lexical du RL-fr (section 2) et justification de l'approche adoptée pour organiser l'information lexicale à travers l'examen de son potentiel applicatif (section 3).

Les systèmes lexicaux appartiennent à la grande famille des réseaux lexicaux, dont la branche issue de WordNet (Fellbaum, 1998) est de loin la plus connue en Traitement Automatique des Langues. Dans ce contexte, il est sans doute utile de rappeler, avant d'entrer dans le vif du sujet, que l'on peut distinguer deux grands types de réseaux lexicaux.

1. Les réseaux ontologiquement organisés, comme WordNet, structurent l'information lexicale de façon prioritairement hiérarchique, en se référant à une organisation de la connaissance en termes de classes et sous-classes. De tels réseaux visent plus particulièrement la modélisation de l'interface monde-lexique à travers la mise en relation de l'univers (profond) des concepts avec l'univers (plus superficiel) des unités lexicales.

*. Cet article est dédié à Anaïs Ferté ; cela fait tellement de mal et tellement de bien, *anafe*, de te croiser au détour du réseau lexical.

2. Les réseaux non ontologiques privilégient quant à eux la modélisation du caractère relationnel des lexiques, en ignorant ou en mettant au second plan la problématique de la classification. Outre les systèmes lexicaux – dont il va être question ici et qui sont manuellement construits –, il s’agit généralement de grands graphes de relations interlexicales (synonymie, inclusion sémantique, etc.) extraits automatiquement des dictionnaires (Ploux & Victorri, 1998) ou d’autres ressources lexicales préexistantes (Spohr, 2012).

2 Le système lexical du Réseau Lexical du Français (RL-fr)

Un système lexical (angl. *lexical system*) d’une langue \mathcal{L} est un modèle du lexique de \mathcal{L} qui possède certaines caractéristiques bien spécifiques de contenu informationnel et de structuration. La notion de système lexical a été théorisée et expérimentée pour la première fois à partir des données lexicographiques de la base lexicale du français DiCo (Polguère, 2000; Mel’čuk & Polguère, 2006) et une description préliminaire du modèle a été publiée dans (Polguère, 2009).

Au moment où a été élaborée la proposition initiale, il n’existait pas d’implémentation de système lexical grandeur nature permettant d’explicitier concrètement le type de modèle postulé et seul un échantillon avait été produit à fin d’illustration. Le RL-fr est la première implantation véritable d’un système lexical et il est désormais possible de s’appuyer sur les résultats obtenus lors de la première étape de construction du RL-fr pour affiner et préciser la notion de système lexical.

Les systèmes lexicaux possèdent quatre caractéristiques principales qui, prises ensemble, les distinguent des autres modèles lexicaux. Elles concernent : les éléments constitutifs du système lexical en tant que graphe – nœuds et arcs (2.1) ; le caractère non atomique de ses nœuds lexicaux (2.2) ; sa structure non ontologique de type graphe petit monde (2.3) ; sa modélisation relativiste de l’information lexicale (2.4).

2.1 Éléments constitutifs

Le système lexical de la langue \mathcal{L} est, formellement, un graphe orienté. Les nœuds de ce graphe correspondent aux différentes entités lexicales de \mathcal{L} . Ces entités sont principalement les unités lexicales proprement dites de \mathcal{L} , c’est-à-dire :

- les lexèmes de \mathcal{L} – TORCHON I [*essuyer avec un torchon*], TORCHON II [*Cet article est un torchon.*]... ;
- les locutions de \mathcal{L} – « COUP DE TORCHON », « MÉLANGER LES TORCHONS ET LES SERVIETTES »...¹

Il peut aussi s’agir, plus marginalement, d’expressions phraséologiques non lexicalisées, notamment de clichés linguistiques comme *Chien méchant* ou *Je vous en prie*².

Dans tous les cas, il est fondamental que chaque nœud du graphe soit associé à un sens de \mathcal{L} bien spécifique. Il ne peut s’agir notamment d’un vocable polysémique (grosso modo, une entrée de dictionnaire), qui est dans les faits un regroupement d’unités lexicales (acceptions du vocable). Ainsi, le vocable TORCHON correspond dans le système lexical à un ensemble regroupant deux nœuds lexicaux {TORCHON I, TORCHON II}. Un système lexical est donc un graphe « désambiguïsé », si on le compare, par exemple, au graphe de synonymie DicoSyn présenté dans (Ploux, 1997). Les nœuds de ce dernier graphe sont en quelque sorte une compaction de copolysèmes et correspondent aux vocables de la langue³. Noter que nous désignerons dorénavant les unités lexicales (lexèmes ou locutions) par le terme plus compact de *lexie*.

Les arcs du système lexical correspondent, dans leur immense majorité, à des liens de fonctions lexicales Sens-Texte (Mel’čuk, 1996). Rappelons que le système des fonctions lexicales dites *standard* (Polguère, 2007) permet d’encoder formellement les relations lexicales paradigmatisées de dérivation sémantique – synonymie, antonymie, conversivité, noms d’actants, adjectifs actanciels... – et les relations syntagmatiques de cooccurrence collocationnelle – intensificateurs, verbes supports... Ainsi, la fonction lexicale **Oper**₁ correspond aux verbes supports qui prennent le premier actant d’une lexie comme sujet et la lexie elle-même comme complément, tel qu’illustré en (1) ci-dessous avec la lexie HOLD-UP I, qui est la lexie de base (non métaphorique) du vocable HOLD-UP.

1. Les noms d’unités lexicales sont écrits en petites majuscules, éventuellement accompagnés de numéros distinctifs en gras. Les noms d’unités lexicales qui sont des locutions sont encadrés par les symboles « \ulcorner \urcorner ».

2. Pour une présentation détaillée de la terminologie lexicale utilisée, on peut se reporter à (Mel’čuk *et al.*, 1995).

3. Fort naturellement, cette approche considère comme unité lexicale ce que nous appelons *vocable*. Il est alors possible à S. Ploux et B. Victorri de parler d’*unités lexicales polysémiques* (Ploux & Victorri, 1998). Dans notre terminologie, une telle expression serait antinomique puisque notre unité de description du lexique – l’unité lexicale – est par définition associée à un seul et unique sens.

(1) **Oper₁**(HOLD-UP I) = *commettre* [ART ~], *faire II.1* [ART ~], *réaliser III* [ART ~]⁴

La Figure 1 est la traduction sous forme de micrographe lexical de l'application de fonction lexicale (1).

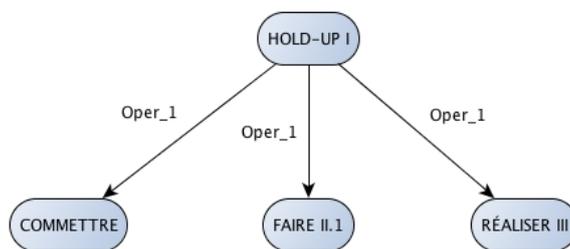


FIGURE 1 – Micrographe représentant **Oper₁**(HOLD-UP I)

Attention cependant ! La Figure 1 correspond à la modélisation de (1) selon les principes de structuration des systèmes lexicaux. Sur le plan de l'informatisation de ces derniers, cependant, rien ne nous oblige à implanter directement de tels graphes. Ainsi, tel que décrit dans (Gader *et al.*, 2012), le système lexical du RL-fr est implanté sous la forme d'une base SQL extrêmement modulaire permettant de gérer efficacement le processus complexe d'édition lexicographique. Nous fonctionnons alors par associations entre entités élémentaires de la base. Dans le cas des liens de fonctions lexicales, une application de fonction telle que celle de (1) est modélisée comme : une association entre l'ID d'une lexie donnée (HOLD-UP I) avec l'ID d'une fonction lexicale donnée (**Oper₁**), association dont l'ID pointe elle-même vers les ID des trois lexies cibles de l'application de la fonction lexicale (COMMETTRE, FAIRE II.1 et RÉALISER III). Trois tables SQL distinctes sont principalement mises à l'œuvre pour encoder ces connexions : (i) *ln_senses*, qui est la table des entités lexicales nœuds du graphe, (ii) *ln_lf*, qui est la table des fonctions lexicales du RL-fr et (iii) *ln_senses_lf*, qui est la table des associations entre les ID des entités de *ln_senses* et les ID des entités de *ln_lf*⁵.

Un tel encodage se visualise en terme de graphe par la Figure 2 ci-dessous, ce qui correspond à une structure radicalement distincte de celle de la Figure 1.

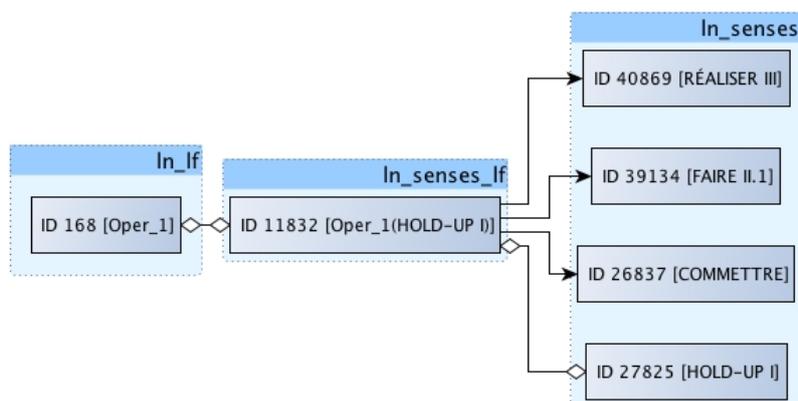


FIGURE 2 – Implantation dans le RL-fr du micrographe de la Figure 1

4. Les numéros distinctifs utilisés pour identifier les sens lexicaux dans cet article correspondent à ceux en vigueur dans le RL-fr au moment où nous écrivons. Ainsi, l'absence de numérotation après *commettre* découle du fait que la polysémie du vocable COMMETTRE n'est pas encore dégagée et qu'à l'heure actuelle seule existe dans le RL-fr l'acception correspondant à l'emploi en tant que verbe support de type **Oper₁** (*commettre un hold-up, un péché, une erreur...*).

5. En d'autres termes, il s'agit de la table qui modélise les applications de fonctions lexicales à des lexies spécifiques. Dans les noms de tables SQL, *ln* est mis pour *lexical network* et *lf* pour *lexical function*.

Cet article vise avant tout la présentation des systèmes lexicaux en tant que « théorisation » de la structure des lexiques, à partir des acquis du travail de construction du RL-fr. Dans la pratique cependant, il convient de choisir le type de structure informatique le plus approprié à l'usage que l'on compte faire des données du réseau lexical. Comme mentionné ci-dessus, nous avons utilisé une structure de type SQL pour répondre aux besoins particuliers de l'édition lexicographique : opérations de création/destruction de nœuds, tissage/détissage de liens, ajout d'informations riches associées à chaque nœud, etc., qui sont effectuées par une équipe complète de lexicographes travaillant simultanément sur les mêmes données. Des structures XML, RDF, etc., peuvent être préférées pour d'autres usages⁶.

Les liens paradigmatiques et syntagmatiques de fonctions lexicales Sens-Texte forment l'ossature principale des systèmes lexicaux, mais ils ne sont pas les seuls à tisser les relations lexicales de la langue. On trouve ainsi notamment au sein des systèmes lexicaux les trois types de liens suivants entre nœuds du graphe :

1. liens de copolysémie – par exemple, un lien métaphore : comme si connecte la lexie HOLD-UP II [*Ce vote est un hold-up démocratique.*] à la lexie de base HOLD-UP I ;
2. liens asémantiques entre les locutions et les lexies dont elles sont formellement constituées – par exemple, la locution 「DONNER CORPS」 [*donner corps à une idée*] pointe vers les lexies DONNER III.2 et CORPS I.1a ;
3. liens d'inclusion sémantique définitionnelle – par exemple la lexie HOLD-UP I contient dans sa définition le sens de la lexie VOLER² I.

Pour l'instant, seuls les deux premiers types de liens additionnels énumérés ci-dessus – copolysémie et inclusion formelle dans les locutions – sont implantés dans le RL-fr. Le travail d'encodage des liens d'inclusion sémantique définitionnelle vient à peine de débuter.

Comme on le voit à l'examen des différents profils de nœuds et arcs présents dans un système lexical, ce modèle a notamment la particularité de correspondre à une structure de graphe relativement hétérogène. Nous considérons que cette hétérogénéité, si elle peut être vue comme un handicap sur le plan formel, n'est en fait que le reflet d'une propriété intrinsèque de l'organisation du lexique des langues.

Pour conclure ce premier niveau de caractérisation des systèmes lexicaux, nous résumons dans le Tableau 1 ci-dessous les statistiques actuelles du RL-fr en termes de nombre de nœuds et d'arcs constitutifs du graphe.

Nombre de nœuds	22 824
Nombre d'arcs de liens de fonctions lexicales	40 361
Nombre d'arcs de copolysémie	2 898
Nombre d'arcs « locution→lexies formellement incluses »	5 028
Degré moyen des nœuds (taux de connectivité)	2,12

Tableau 1 – Statistiques actuelles du RL-fr en termes de nœuds et d'arcs

Il convient de noter que le tissage des liens de copolysémie a débuté récemment, ce qui explique le fait que très peu d'entre eux soient pour l'instant présents dans le RL-fr. On remarque aussi dans les statistiques du Tableau 1 que la taille du RL-fr commence à être significative, en termes d'entités lexicales décrites (principalement, des lexies), mais que le taux de connectivité est encore bas : certaines lexies bien décrites sont fortement connectées, alors que d'autres sont presque isolées dans le graphe. Le taux de connectivité croît cependant de façon constante.

2.2 Caractère non atomique des nœuds lexicaux

Dans un système lexical, les nœuds lexicaux du graphe ne sont pas atomiques, en ce sens qu'ils possèdent eux-mêmes une structure interne complexe. « Ouvrir » un nœud lexical pour regarder ce qu'il y a à l'intérieur revient en quelque sorte à afficher l'article lexicographique correspondant. Chaque nœud représentant une lexie est ainsi associé à un ensemble structuré de propriétés, qui sont principalement :

- les caractéristiques grammaticales de la lexie – parties du discours, genre (pour les noms), tables de flexion, etc. ;
- son sens – étiquette sémantique (Polguère, 2011), forme propositionnelle (c.-à-d., structure actancielle) et définition proprement dite ;

6. Ainsi, les visualisations présentées dans la section 3.2 ci-dessous sont générées à partir de graphes construits en Python à l'aide de la bibliothèque `igraph`. (Merci à Yann Desalle pour ces informations.)

- son régime syntaxique ;
- ses connexions paradigmatiques et syntagmatiques – directement implantées par le tissage des liens du graphe lexical ;
- des illustrations de ses emplois – exemples lexicographiques.

Ces ensembles de propriétés correspondent aux différentes zones d'un article de lexicographique d'un *Dictionnaire Explicatif et Combinatoire* (Mel'čuk et al., 1984 1988 1992 1999) ou de toute base lexicale conçue selon les principes de la Lexicologie Explicative et Combinatoire. La Figure 3 illustre ce fait en présentant un extrait de la *vue-article* de HOLD-UP I dans son état courant, telle que générée par l'éditeur lexicographique Dicet (Gader *et al.*, 2012)⁷.

hold-up I	<p>Caractéristiques grammaticales</p> <p>angl nom commun masc invar graphie rectifiée : « holdup »</p> <p>Définition</p> <p>méfait</p> <p>hold-up de X=1 contre Y=2</p> <p>attaque I₂ par l'individu X du lieu public Y au moyen d'une arme dans le but d'y commettre un voF I</p> <p>Fonctions lexicales</p> <p>Syn, attaque I₂</p> <p>Syn, voF I; fam braquage; fam casse</p> <p>Nom pour X</p> <p>S₁, voleur, I, voleuse I; bandit I</p> <p>Type particulier de Y</p> <p>type particulier S₂ banque; bijouterie; magasin</p> <p>[X] commettre ~</p> <p>Oper, commettre [ART ~], faire II.1 [ART ~], réaliser III [ART ~]</p> <p>Ce que X dit quand il commet ~</p> <p>Enun, Real, Haut les mains I, Les mains en l'air !</p> <p>AntiBonFact, tourner mal</p> <p>Exemples</p> <p>Oui, il a lu le dossier, cet énorme pavé (que j'ai moi-même soigneusement parcouru, dans ma cellule, des jours entiers), où tous les témoignages, les interrogatoires, la reconstitution du hold-up ont été consignés, il y a plus de vingt ans.</p> <p>Frantext CASTEL Hélène, <i>Retour d'exil d'une femme recherchée</i>, 2009, p. 153</p> <p>Il est par contre exact que j'ignorais (et ignore encore) l'identité des malfaiteurs qui m'ont accompagné lors des hold-up commis contre la société Vog et le payeur des Allocations familiales du passage Ramey.</p> <p>Frantext GOLDMAN Pierre, <i>Souvenirs obscurs d'un juif polonais né en France</i>, 1975, p. 258</p>
-----------	--

FIGURE 3 – Extrait de la vue-article de HOLD-UP I dans l'éditeur Dicet

La vue-article d'une lexie est la rétroaction de forme textuelle que l'éditeur Dicet génère pour le lexicographe à partir des informations linguistiques associées à la lexie en question. Il ne s'agit donc aucunement d'un texte éditable par le lexicographe. Ce dernier n'écrit pas d'article lexicographique et son activité principale consiste à « tisser sa toile » (Polguère, 2012b). Cette approche lexicographique – suivant laquelle le texte dictionnaire n'est pas écrit, mais est virtuellement encapsulé dans le modèle lexical (Atkins, 1996; Polguère, 2012a) – nous semble celle qui se prête le mieux à la construction des systèmes lexicaux du fait de la nature fondamentalement non textuelle de ces derniers.

Mentionnons, pour conclure cette section, que les propriétés « internes » des nœuds lexicaux sont, comme leurs propriétés relationnelles, tout à fait modélisables sous forme de graphe. Dans le cas du RL-fr, chaque propriété associée à un nœud lexical est implantée comme un pointeur vers une entité autonome de la base. Indiquer qu'une lexie est un nom commun, est du genre masculin ou est familière revient à tisser un lien entre la lexie en question et des entités métalinguistiques stockées dans des modèles intégrés des caractéristiques grammaticales, des étiquettes sémantiques, des fonctions lexicales, etc. Le RL-fr pousse en fait cette approche à son extrême, puisque même les exemples lexicographiques y sont des entités autonomes vers lesquelles pointent les lexies individuelles⁸. Plusieurs lexies peuvent ainsi pointer vers le même exemple,

7. Le régime syntaxique (structure de complémentation) n'apparaît pas pour l'instant dans la vue-article.

8. L'ensemble des citations utilisées dans le RL-fr constitue en soi un corpus linguistique enchâssé dans le modèle. Au moment où nous écrivons, le corpus des citations du RL-fr totalise plus d'un million de mots-occurrences ; il est notamment accessible par les lexicographes au moyen de concordances effectuées avec le logiciel TXM (Heiden *et al.*, 2010). Le corpus enchâssé des citations ne doit bien entendu pas être confondu avec les différents corpus de référence dont les lexicographes font usage et dont les citations du RL-fr sont justement extraites.

chaque pointeur visant une séquence linguistique spécifique dans l'exemple ; cela permet entre autres d'ouvrir la « fiche lexicographique » d'un exemple donné et d'accéder à partir d'elle à toutes les lexies qui y font appel. C'est ce qu'illustre la Figure 4 ci-dessous, qui est la visualisation sous Dicet de l'information associée au second exemple lexicographique apparaissant dans la Figure 3 (article de HOLD-UP I).

ID : 470
 Classe : Citations de corpus type Frantext
 Source : Frantext
 Statut : En attente de validation (1)
 Citation
 Il est par contre exact que j'ignorais (et ignore encore) l'identité des malfaiteurs qui m'ont accompagné lors des hold-up commis contre la société Vog et le payeur des Allocations familiales du passage Ramey.

Référence
 Frantext GOLDMAN Pierre, *Souvenirs obscurs d'un juif polonais né en France*, 1975, p. 258

Occurrences

hold-up I	Il est par contre exact que j'ignorais (et ignore encore) l'identité des malfaiteurs qui m'ont accompagné lors des hold-up commis contre la société Vog et le payeur des Allocations familiales du passage Ramey.
malfaiteur	Il est par contre exact que j'ignorais (et ignore encore) l'identité des malfaiteurs qui m'ont accompagné lors des hold-up commis contre la société Vog et le payeur des Allocations familiales du passage Ramey.

Création par debe le 8 janvier 2013 à 09:16:09
 Modification par mvs le 23 janvier 2013 à 17:53:59

FIGURE 4 – Fiche d'un exemple lexicographique dans l'éditeur Dicet

2.3 Structure non ontologique de type petit monde

Comme mentionné dans l'Introduction, les systèmes lexicaux sont des modèles non ontologiques. Cela découle directement du fait que l'ossature de ces modèles est tissée à partir du système des fonctions lexicales : il s'agit d'un système de relations sémantico-syntaxiques qui se distingue radicalement de l'organisation hiérarchique des concepts, en classes et sous-classes, postulées par les modèles ontologiques. Pour dire les choses de façon caricaturale : la visualisation d'un lexique ontologique évoque une sorte de tour Eiffel aplatie, alors que celle d'un système lexical présente plutôt une analogie avec un plat de spaghetti.

Le graphe d'un système lexical peut ainsi sembler être un « fouillis » de relations paradigmatiques et syntagmatiques où n'émerge aucune structure classifiante. Pourtant, un tel graphe possède au moins quatre propriétés formelles remarquables⁹ :

- P1** La densité en arcs est faible : les graphes lexicaux, et les systèmes lexicaux en particulier, possèdent peu d'arcs, en regard de leur nombre total de sommets.
- P2** La moyenne des plus courts chemins est petite : en général, il existe au moins un chemin relativement court entre n'importe quelle paire de sommets.
- P3** Il existe des *clusters* : les graphes lexicaux sont localement denses en arcs, alors que leur densité globale est faible (propriété **P1**).
- P4** La distribution du degré d'incidence des sommets suit une loi de puissance : la probabilité qu'un sommet donné ait k voisins décroît comme une loi de puissance de k (alors que c'est une loi de Poisson dans les graphes aléatoires).

Les graphes possédant les propriétés **P1**, **P2** et **P3** sont appelés *graphes petits mondes* (angl. *small-world networks/graphs*), et ont été mathématiquement définis dans (Watts & Strogatz, 1998).

9. La formulation qui suit nous a été suggérée par Bruno Gaume, que nous remercions pour nous avoir aidé à mettre plus de rigueur mathématique dans nos propos.

Les propriétés des petits mondes lexicaux ont été systématiquement étudiées par Bruno Gaume (Gaume, 2004; Gaume *et al.*, 2006; Gaume, 2008), qui a démontré qu'il était possible d'exploiter leurs particularités formelles pour y identifier des espaces sémantiques structurés et, de façon plus générale, pour mettre de l'ordre dans le désordre lexical apparent.

Le RL-fr, même s'il s'agit d'un réseau lexical encore peu développé, est déjà suffisamment riche pour manifester les caractéristiques des graphes petits mondes (Gader *et al.*, 2014). En théorie, tout système lexical, tel que défini ici, devra posséder ce type de structuration.

2.4 Modèle relativiste

Les systèmes lexicaux incorporent une mesure de flou. Cette propriété des systèmes lexicaux a été anticipée dès la proposition initiale de (Polguère, 2009) pour deux raisons :

1. La connaissance lexicale est elle-même gradable, du fait notamment des changements diachroniques et de l'incertitude linguistique potentielle des locuteurs, y compris des lexicographes.
2. Les systèmes lexicaux doivent être des modèles supportant l'inférence et la construction de la connaissance lexicale par approximation – cf. plus bas, section 3.3.

Le RL-fr implémente cette nature gradable de l'information lexicale. Chaque information linguistique – existence même d'une lexie, ses caractéristiques grammaticales, etc. – est associée à un niveau de confiance. Ce dernier est par défaut maximal lorsque l'information est entrée manuellement par les lexicographes. Ceux-ci peuvent cependant « nuancer » l'information en lui associant un niveau de confiance moindre. Par convention, la confiance minimale qui peut être manuellement entrée est de 60%. Toute information qui sera automatiquement injectée dans la base se verra quant à elle attribuer un niveau de confiance de 50%, et c'est cette confiance spécifique relativement basse, associée aux données résultant d'un enrichissement automatique de la base, qui permet leur identification.

Nous avons ainsi généré de façon automatique un Réseau Lexical de l'Anglais, RL-en, par compilation de l'information lexicographique contenue dans le WordNet de Princeton (Fellbaum, 1998)¹⁰. Toutes les informations du système lexical résultant sont associées à un niveau de confiance de 50% et toute entrée de données qui sera effectuée manuellement sur le RL-en au moyen de l'éditeur Dicet générera par contraste une information validée (niveau de confiance de 100%).

Bien entendu, nous ne faisons pour l'instant qu'un usage un peu caricatural des mesures de confiance associées aux données du RL-fr et du RL-en. Ces ressources sont cependant équipées de tous les dispositifs qui permettront de tirer parti de cette propriété des systèmes lexicaux, notamment pour ce qui est de la semi-automatisation de la croissance de la ressource, sur laquelle nous revenons dans la section 3.3.

3 Justifications de l'approche

Dans cette section, nous tenterons d'apporter plusieurs pistes de justification théorique et empirique pour l'adoption des systèmes lexicaux en tant que modèles formels des lexiques. Nous abordons successivement les trois points suivants : possible intérêt des systèmes lexicaux dans une perspective psycholinguistique (3.1) ; application de l'analyse proxémique (3.2) ; semi-automatisation du processus de croissance du RL-fr et, plus généralement, des systèmes lexicaux (3.3).

3.1 Les systèmes lexicaux comme modèles du lexique mental

On entend par *lexique mental* (angl. *mental lexicon*) la connaissance lexicale telle qu'elle est stockée et structurée dans le cerveau d'un individu (Aitchison, 2003; Zock, 2005; Wierzbicka, 2009). Il existe deux raisons principales pour lesquelles il serait intéressant que les systèmes lexicaux soient une représentation plausible de la structure des lexiques mentaux. Tout d'abord, cela ferait des systèmes lexicaux un support pertinent pour mener des études aussi bien lexicologiques que psycholinguistiques, les modèles construits étant en harmonie avec le « lexique véritable ». Ensuite, on peut faire l'hypothèse que de tels modèles permettraient de simuler informatiquement les tâches d'accès et de manipulation de l'information lexicale avec la même efficacité que le fait le cerveau humain. La recherche d'une adéquation entre modèle formel construit et lexique mental postulé peut donc trouver sa justification aussi bien sur un plan théorique que pratique.

10. La méthode de génération automatique du RL-en est décrite dans (Gader *et al.*, 2014).

On notera que l'adéquation au lexique mental était justement la visée première du projet WordNet de Princeton et que (Miller *et al.*, 1990) se situaient explicitement dans un contexte de *psycholexicologie* (angl. *psycholexicology*). Comme le relève (Zock, 2005, p. 113) cependant, « [...] les psycholinguistes travaillant sur l'accès lexical ne mentionnent pour ainsi dire jamais WordNet. La communauté de WordNet, malgré sa taille, semble complètement ignorer les travaux sur l'accès lexical ». Au-delà de la seule problématique de l'accès lexical, c'est toute la perspective psycholinguistique qui est quasi absente des travaux menés sur et à partir de WordNet¹¹. Ces travaux exploitent dans leur grande majorité la très haute qualité lexicographique de WordNet (plutôt que sa validité psycholinguistique postulée) dans un contexte de Traitement Automatique de la Langue.

Tout cela ne veut bien entendu pas dire que WordNet n'est pas psycholinguistiquement pertinent. Simplement, sa pertinence sur ce plan est loin d'être démontrée et il est, par conséquent, tout à fait légitime d'explorer d'autres avenues. Nous postulons quant à nous, sans pouvoir le démontrer pour l'instant et sans même prétendre nous appuyer sur des acquis de la psychologie moderne¹², que l'approche des systèmes lexicaux est compatible avec des exploitations psycholinguistiques. Nous voyons les systèmes lexicaux comme une approximation de la structure des lexiques mentaux, si l'on entend par ce terme spécifiquement la connaissance linguistique lexicale. Nous ne sommes bien entendu aucunement en mesure d'émettre quelque hypothèse que ce soit sur la façon dont les concepts eux-mêmes et la connaissance extra-linguistique sont mentalement organisés.

À l'appui du recours aux systèmes lexicaux pour l'étude du lexique mental, nous nous contenterons de mentionner la compatibilité formelle de ces derniers avec de nombreuses propositions faites pour la modélisation de l'acquisition du vocabulaire en tant qu'enrichissement de réseaux lexicaux – voir, par exemple, (Wolter, 2006) et (Gaume *et al.*, 2008). Les domaines d'application qui méritent d'être explorés prioritairement pour valider l'hypothèse d'une pertinence psycholinguistique des systèmes lexicaux nous semblent être : l'analyse des erreurs lexicales en parole spontanée et le diagnostic de troubles mentaux ou physiologiques se manifestant sur le plan de l'accès lexical.

3.2 Identification d'espaces sémantiques dans les systèmes lexicaux

Nous entendons ici par *espace sémantique* d'un réseau lexical représentant le lexique d'une langue \mathcal{L} un ensemble plutôt petit d'entités lexicales de \mathcal{L} présentant une forme de cohérence sémantique. La notion d'espace sémantique est proche de celle de champ sémantique, mais elle s'en distingue par le fait que tous les éléments d'un tel espace n'appartiennent pas nécessairement à un champ sémantique donné¹³. En effet, le sens d'une lexie est directement connecté à sa combinatoire et les lexies sémantiquement proches tendent à partager de façon significative des éléments de combinatoire. Cela vaut bien entendu pour ce que l'on appelle la *combinatoire libre*, qui découle directement de la définition de la lexie, mais aussi pour la combinatoire lexicale restreinte. Ainsi, même si l'on doit considérer qu'un syntagme comme *commettre un hold-up* – analysé plus haut, section 2.1 – est une collocation de type **Oper₁** contrôlée par la lexie HOLD-UP I, il est logique de postuler que nombre de lexies nominales sémantiquement proches de HOLD-UP I tendront à contrôler le même **Oper₁** : *commettre un crime, un massacre, un meurtre, un vol*, etc. Les lexies d'un même champ sémantique ont donc tendance non seulement à former des « communautés » au sein des graphes lexicaux petits mondes, mais elles ont aussi tendance à exercer une force d'attraction sur des lexies collocationnelles très vagues ou quasi vides et à les aspirer dans leur espace sémantique. Un espace sémantique est donc potentiellement une expansion d'un champ sémantique résultant des affinités lexicales syntagmatiques.

Le fait que les systèmes lexicaux traitent de façon uniforme les liens paradigmatiques et les liens syntagmatiques au sein d'une même structure va permettre de tirer parti de la combinatoire des lexies pour identifier de façon encore plus efficace les espaces sémantiques au sein des lexiques, notamment par le recours à la proxémie. Cette dernière est une mesure de similarité définie sur les graphes petits mondes par B. Gaume (Gaume, 2004) et qui a tout particulièrement été appliquée aux graphes de synonymie (Gaume *et al.*, 2006). Elle permet d'identifier des espaces sémantiques de façon entièrement « non linguistique », c'est-à-dire en s'appuyant exclusivement sur la structure du graphe « nu », exprimée en termes de sommets et d'arcs connectant ces sommets.

L'analyse proxémique par la méthode de parcours Prox (Gaume, 2008) est présentement appliquée sur le RL-fr, sans ajustement particulier, pour en extraire une clusterisation du système lexical à partir de lexies données. Une visualisation par Tmuse (Chudy *et al.*, 2013) permet de représenter graphiquement les résultats obtenus, comme illustré dans la Figure 5 ci-dessous à partir de la lexie FOOTBALL 1 du RL-fr.

11. Ainsi, aucun article présenté au dernier congrès *Global WordNet* (Orav *et al.*, 2014) ne concerne une recherche relevant de la psycholinguistique.

12. Voir néanmoins (Morais *et al.*, 2013).

13. Pour une définition de la notion de champ sémantique, en contraste notamment avec celle de champ lexical, voir (Polguère, 2013).

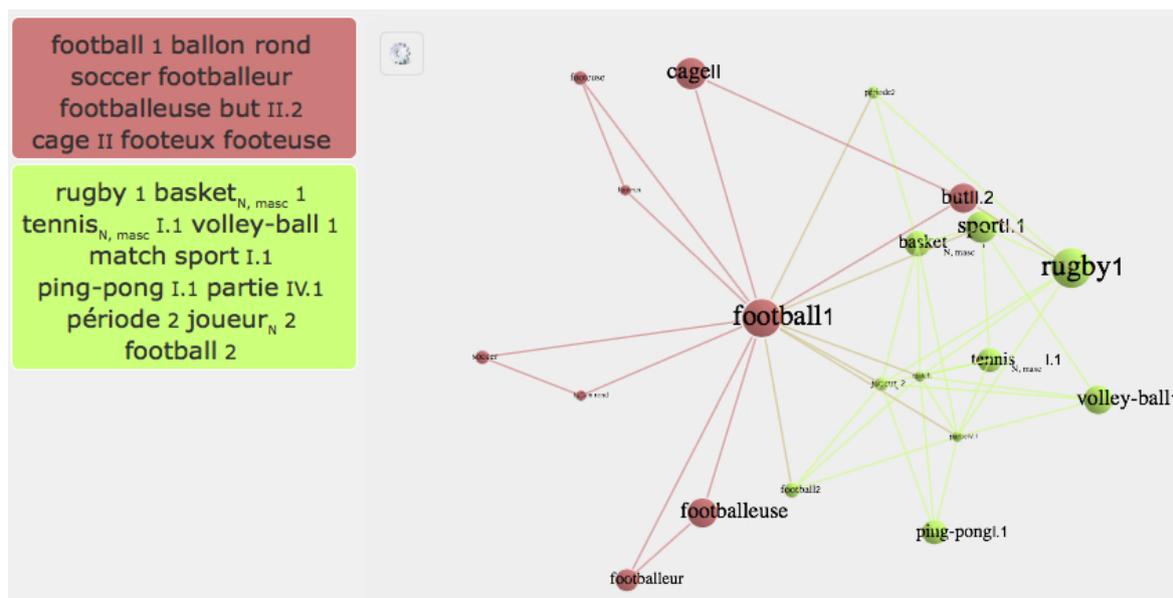


FIGURE 5 – Clusterisation par proximité et visualisation de FOOTBALL 1

On voit que les deux clusters qui ont été ici identifiés par Prox, sans recours à la sémantique même des liens de fonction lexicale, correspondent à une segmentation de l'espace sémantique de la lexie selon une base assez logique, qui est grosso modo : vocabulaire du football vs vocabulaire connecté à celui du football. On notera que FOOTBALL 2, qui apparaît dans le second cluster, aurait en réalité tout à fait sa place dans le premier, puisque cette lexie appartient très directement au vocabulaire du football : FOOTBALL 2 est l'acception métonymique du vocable, que l'on trouve par exemple dans *Tu viens faire un petit football dans la ruelle ?* Cette imperfection dans le résultat obtenu doit possiblement être interprétée comme un indice du fait que la connexion de FOOTBALL 2 au reste du graphe est encore incomplète. Nous n'en sommes qu'au tout début de l'application des analyses proxémiques aux systèmes lexicaux. Il s'agit d'un travail qui va se poursuivre sur le long terme, dans une visée aussi bien applicative que descriptive, comme nous allons maintenant l'expliquer.

3.3 Vers une semi-automatisation de la croissance du RL-fr

Il existe quatre principales méthodes de construction des ressources lexicales destinées au Traitement Automatique de la Langue ou à tout autre domaine d'application langagier :

1. la méthode lexicographique, c'est-à-dire la construction manuelle des ressources lexicales telle qu'elle est pratiquée pour le RL-fr ou telle qu'elle a été pratiquée dès le lancement du projet WordNet (Miller *et al.*, 1990) ;
2. la méthode coopérative, qui repose sur une approche type *crowdsourcing* telle qu'appliquée pour Wiktionary¹⁴ ou, selon une stratégie ludique, avec JeuxDeMots (Lafourcade & Joubert, 2013)¹⁵ ;
3. la méthode automatique à base de ressources lexicales, qui consiste en l'extraction et la structuration informatique de données linguistiques à partir de ressources lexicales préexistantes, comme dans le cas de WOLF (Sagot & Fišer, 2008) ou de BabelNet (Navigli & Ponzetto, 2012) ;
4. la méthode automatique à base de corpus textuels, qui s'appuie exclusivement sur des données linguistiques brutes, et non sur des modélisation métalinguistiques antérieurement construites (Hearst, 1992).

On notera que la troisième méthode présuppose l'activité de linguistes et lexicographes. Sans dictionnaires et autres WordNet, il n'y aurait, par définition, pas de matériau permettant d'alimenter les projets de construction automatique de ressources dérivées de ces ressources mères. Ces ressources mères ne peuvent d'ailleurs jamais être considérées comme

14. <http://www.wiktionary.org>

15. <http://www.jeuxdemots.org/jdm-accueil.php>

ayant été produites une fois pour toutes : la langue évolue et la modélisation lexicographique est dans les faits un travail perpétuellement en cours.

La seconde approche est exclue si l'on veut disposer de ressources proposant une description fine de l'information lexicale et d'une qualité qui permette, par exemple, de l'utiliser dans un contexte pédagogique. De même qu'un patient a des raisons de s'inquiéter lorsqu'il voit son médecin chercher sur Wikipedia de l'information sur une maladie ou un symptôme le concernant, de même un parent d'élève du primaire serait en droit de s'inquiéter de voir l'enseignant de son enfant utiliser Wiktionary, plutôt qu'un dictionnaire élaboré par des lexicographes, pour appuyer son enseignement du vocabulaire.

Il faut donc bien que « quelqu'un s'y colle », et c'est justement la tâche (et le plus grand plaisir) des lexicographes que de produire les ressources lexicales dont, ultimement, tout le monde a besoin.

En dépit de cela, il est plus que légitime d'envisager une part d'automatisation du processus lexicographique. Cela peut aller de l'utilisation d'outils proposant des ébauches de description (Kilgarriff & Rychlý, 2010) à l'enrichissement semi-automatique par inférences fondées sur la structure même de l'information lexicale (Sajous *et al.*, 2011). Il nous semble que la structure des systèmes lexicaux se prête tout particulièrement à ce type de semi-automatisation du travail lexicographique. Les premières observations qui ont été faites sur le traitement automatique du graphe lexical du RL-fr par la proxémie (section 3.2, ci-dessus) sont de ce point de vue encourageantes. La clusterisation permet dans de nombreux cas de mettre en évidence des trous dans l'espace sémantique des lexies, trous résultant de la jeunesse du RL-fr. Il peut s'agir de lexies manquantes dans le modèle ou, beaucoup plus fréquemment, d'une connexion insuffisante au reste du graphe des lexies déjà présentes. Il devient alors intéressant de tenter d'exploiter les algorithmes de parcours de graphes petits mondes pour repérer dans le RL-fr des cliques ou des « presque-cliques »¹⁶ significatives et identifier des micropatrons récurrents de connexions lexicales. Ces patrons pourront ensuite être exploités afin de générer automatiquement des éléments de graphe plausibles¹⁷. Ces derniers, qui devront être validés manuellement, permettront d'obtenir une sorte de « RL-fr turbo » constitué d'un noyau entièrement validé, pour l'exploitation humaine, et d'une extension à plus faible taux de confiance, pour compléter le RL-fr noyau dans un contexte de Traitement Automatique de la Langue notamment.

4 Conclusion

On ne peut bien entendu pas considérer la pertinence des systèmes lexicaux comme étant un fait acquis, tant que l'efficacité de ces modèles n'aura pas été démontrée à travers leur mise en pratique. Pour cela, il va falloir expérimenter, notamment dans les domaines de la recherche d'information, de la didactique des langues ou de l'étude psycholinguistique. Cependant, il est d'ores et déjà justifié de dire qu'une première étape essentielle dans la validation de l'approche des systèmes lexicaux a été franchie. Il a en effet été démontré, à travers le travail de construction du RL-fr, que le mode de structuration offert par les systèmes lexicaux permet de mener avec efficacité un programme lexicographique véritablement conçu comme le tissage d'un graphe lexical¹⁸.

Remerciements

Le projet RELIEF, auquel participe la société MVS de Sainte-Marguerite, est financé par l'Agence de Mobilisation Économique de la Région Lorraine et le FEDER Lorrain. Les résultats présentés ici ont été obtenus grâce au travail considérable effectué par les lexicographes du RL-fr (<http://www.atilf.fr/spip.php?article908>) et par Nabil Gader, développeur de l'éditeur lexicographique Dicot; nous les remercions tous chaleureusement. Le travail en cours sur la visualisation du graphe du RL-fr et la clusterisation par proxémie est effectué par Yann Desalle, en collaboration avec Bruno Gaume et la Proxteam de Toulouse (<http://www.irit.fr/Proxteam/>). Merci à Bruno Gaume, Sandrine Ollinger et à trois relecteurs anonymes de TALN 2014 pour leurs commentaires sur une première version de cet article.

16. Nous entendons par là des sous-graphes qui ne sont pas strictement parlant des cliques (sous-graphes complets), mais dont on peut considérer que le nombre maximal d'arcs est atteint en regard de contraintes spécifiques associées aux données que le graphe modélise. Par exemple, tel sous-graphe d'un système lexical n'est pas une clique, car les deux sommets s_1 et s_2 ne sont pas liés; **mais** on peut considérer que c'est une presque-clique, car les deux unités lexicales représentées par ces sommets ne devraient de toute façon pas être liées par des liens paradigmatiques ou syntagmatiques, pour des raisons linguistiques.

17. Cf. la recherche doctorale de S. Ollinger, en cours à l'ATILF : *Le raisonnement analogique en lexicographie et son informatisation : application au Réseau Lexical du Français*.

18. Le graphe lexical du RL-fr doit être rendu disponible par téléchargement libre d'ici la fin du projet RELIEF (automne 2014), ce qui doit représenter la fin de première étape dans construction de la ressource. D'ici là, il est bien entendu possible à tout chercheur désireux de conduire des expérimentations avec les données du RL-fr de nous contacter directement.

Références

- AITCHISON J. (2003). *Words in the Mind : An Introduction to the Mental Lexicon*. Oxford, G.-B. : Blackwell.
- ATKINS B. T. S. (1996). Bilingual Dictionaries : Past, Present and Future. In M. GELLERSTAM, J. JÄRBORG, S.-G. MALMGREN, K. NORÉN, L. ROGSTRÖM & C. R. PAPMEHL, Eds., *Euralex'96 Proceedings*, p. 515–590, Gothenburg : Gothenburg University, Department of Swedish.
- CHUDY Y., DESALLE Y., GAILLARD B., GAUME B., MAGISTRY P. & NAVARRO E. (2013). Tmuse : Lexical Network Exploration. In *The Companion Volume of the Proceedings of IJCNLP 2013 : System Demonstrations*, p. 41–44, Nagoya, Japan : Asian Federation of NLP.
- C. FELLBAUM, Ed. (1998). *WordNet : An Electronic Lexical Database*. Cambridge, MA : The MIT Press.
- GADER N., LUX-POGODALLA V. & POLGUÈRE A. (2012). Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor. In *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex III)*, p. 109–125, Mumbai : The COLING 2012 Organizing Committee.
- GADER N., OLLINGER S. & POLGUÈRE A. (2014). One lexicon, two structures : So what gives ? In H. ORAV, C. FELLBAUM & P. VOSSEN, Eds., *Proceedings of the Seventh Global Wordnet Conference (GWC2014)*, p. 163–171, Tartu : Global WordNet Association.
- GAUME B. (2004). Balades aléatoires dans les Petits Mondes Lexicaux. *13 Information Interaction Intelligence*, **4**(2), 39–96. CEPADUES édition.
- GAUME B. (2008). Mapping the Forms of Meaning in Small Worlds. *Journal of Intelligent Systems*, **23**, 848–862.
- GAUME B., DUVIGNAU K., PRÉVOT L. & DESALLE Y. (2008). Toward a cognitive organization for electronic dictionaries, the case for semantic proxemy. In *Proceedings of the workshop on Cognitive Aspects of the Lexicon (COGALEX 2008)*, p. 86–93, Manchester.
- GAUME B., VENANT F. & VICTORRI B. (2006). Hierarchy in Lexical Organization of Natural Languages. In D. PUMAIN, Ed., *Hierarchy in Natural and Social Sciences*, Methodos series 3, p. 121–142. Dordrecht : Springer.
- HEARST M. A. (1992). Automatic Acquisition of Hyponyms from Large Text Corpora. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING 1992)*, p. 539–545, Nantes.
- HEIDEN S., MAGUÉ J.-P. & PINCEMIN B. (2010). TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. In S. BOLASCO, I. CHIARI & L. GIULIANO, Eds., *Statistical Analysis of Textual Data – Proceedings of 10th International Conference Journées d'Analyse statistique des Données Textuelles*, volume 2-3, p. 1021–1032, Rome : Edizioni Universitarie di Lettere Economia Diritto.
- KILGARRIFF A. & RYCHLÝ P. (2010). Semi-automatic Dictionary Drafting. In G.-M. DE SCHRYVER, Ed., *A Way with Words : Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*, p. 299–312. Kampala : Menha Publishers.
- LAFOURCADE M. & JOUBERT A. (2013). Bénéfices et limites de l'acquisition lexicale dans l'expérience jeuxdemots. In N. GALA & M. ZOCK, Eds., *Ressources Lexicales : Contenu, construction, utilisation, évaluation*, *LinguisticæInvestigationes*, Supplementa 30, p. 187–216. Amsterdam/Philadelphia : John Benjamins.
- LUX-POGODALLA V. & POLGUÈRE A. (2011). Construction of a French Lexical Network : Methodological Issues. In *Proceedings of the First International Workshop on Lexical Resources, WoLeR 2011. An ESSLLI 2011 Workshop*, p. 54–61, Ljubljana.
- MEL'ČUK I. (1996). Lexical Functions : A Tool for the Description of Lexical Relations in the Lexicon. In L. WANNER, Ed., *Lexical Functions in Lexicography and Natural Language Processing*, volume 31 of *Language Companion Series*, p. 37–102. Amsterdam/Philadelphia : John Benjamins.
- MEL'ČUK I. (2006). Explanatory Combinatorial Dictionary. In G. SICA, Ed., *Open Problems in Linguistics and Lexicography*, p. 225–355. Monza : Polimetrica.
- MEL'ČUK I., CLAS A. & POLGUÈRE A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Paris/Louvain-la-Neuve : Duculot.
- MEL'ČUK I. & POLGUÈRE A. (2006). Dérivations sémantiques et collocations dans le DiCo/LAF. *Langue française*, **150**, 66–83.
- MEL'ČUK ET AL. I. (1984, 1988, 1992, 1999). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques. Volumes I–IV*. Montréal : Les Presses de l'Université de Montréal.

- MILLER G. A., BECKWITH R., FELLBAUM C., GROSS D. & MILLER K. J. (1990). Introduction to WordNet : An On-line Lexical Database. *International Journal of Lexicography*, **3**(4), 235–244.
- MORAIS A. S., OLSSON H. & SCHOOLER L. J. (2013). Mapping the Structure of Semantic Memory. *Cognitive Science*, **37**, 125–145.
- NAVIGLI R. & PONZETTO S. P. (2012). BabelNet : The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. *Artificial Intelligence*, **193**, 217–250.
- H. ORAV, C. FELLBAUM & P. VOSSEN, Eds. (2014). *Proceedings of the Seventh Global Wordnet Conference (GWC2014)*, Tartu (Estonia). Global WordNet Association.
- PLOUX S. (1997). Modélisation et traitement informatique de la synonymie. *Linguisticæ Investigationes*, **21**(1), 1–27.
- PLOUX S. & VICTORRI B. (1998). Construction d’espaces sémantiques à l’aide de dictionnaires de synonymes. *Traitement Automatique des Langues (T.A.L.)*, **39**(1), 161–182.
- POLGUÈRE A. (2000). Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. In *Proceedings of EURALEX’2000*, p. 517–527, Stuttgart.
- POLGUÈRE A. (2007). Lexical function standardness. In L. WANNER, Ed., *Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In Honour of Igor Mel’čuk*, volume 84 of *Language Companion Series*, p. 43–95. Amsterdam/Philadelphia : John Benjamins.
- POLGUÈRE A. (2009). Lexical systems : graph models of natural language lexicons. *Language Resources and Evaluation*, **43**(1), 41–55.
- POLGUÈRE A. (2011). Classification sémantique des lexies fondée sur le paraphrasage. *Cahiers de lexicologie*, **98**, 197–211.
- POLGUÈRE A. (2012a). Lexicographie des dictionnaires virtuels. In Y. APRESJAN, I. BOGUSLAVSKY, M.-C. L’HOMME, L. IOMDIN, J. MILIČEVIĆ, A. POLGUÈRE & L. WANNER, Eds., *Meanings, Texts, and Other Exciting Things. A Festschrift to Commemorate the 80th Anniversary of Professor Igor Alexandrovič Mel’čuk*, *Studia Philologica*, p. 509–523. Moscou : Jazyki slavjanskoj kultury Publishers.
- POLGUÈRE A. (2012b). Like a Lexicographer Weaving Her Lexical Network. In *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex III)*, p. 1–3, Mumbai : The COLING 2012 Organizing Committee.
- POLGUÈRE A. (2013). Les petits soucis ne poussent plus dans le champ lexical des sentiments. In F. BAIDER & G. CISLARU, Eds., *Cartographie des émotions. Propositions linguistiques et sociolinguistiques*, p. 21–41. Paris : Presses Sorbonne Nouvelle.
- POLGUÈRE A. & SIKORA D. (2013). Modèle lexicographique de croissance du vocabulaire fondé sur un processus aléatoire, mais systématique. In C. GARCIA-DEBANC, C. MASSERON & C. RONVEAUX, Eds., *Enseigner le lexique, Recherches en didactique du français 5*, p. 35–63. Namur : Presses Universitaires de Namur.
- SAGOT B. & FIŠER D. (2008). Building a free French wordnet from multilingual resources. In *Proceedings of OntoLex 2008*, Marrakech.
- SAJOUS F., NAVARRO E. & GAUME B. (2011). Enrichissement de lexiques sémantiques approvisionnés par les foules : le système WISIGOTH appliqué à Wiktionary. *Traitement Automatique des Langues (T.A.L.)*, **52**(1), 11–35.
- SPOHR D. (2012). *Towards a Multifunctional Lexical Resource. Design and Implementation of a Graph-based Lexicon Model*. Berlin/Boston : De Gruyter.
- WATTS D. J. & STROGATZ S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, **393**, 440–442.
- WIERZBICKA A. (2009). The theory of the mental lexicon. In S. KEMPGEN, P. KOSTA, T. BERGER & K. GUTSCHMIDT, Eds., *Die slavischen Sprachen/The Slavic Languages : An International Handbook of their Structure, their History and their Investigation*, p. 848–863. Berlin & New York : Mouton de Gruyter.
- WOLTER B. (2006). Lexical Network Structures and L2 Vocabulary Acquisition : The Role of L1 Lexical/Conceptual Knowledge. *Applied Linguistics*, **27**(4), 741–747.
- ZOCK M. (2005). Le dictionnaire mental, modèle des dictionnaires de demain ? *Revue française de linguistique appliquée*, **X**(2), 103–117.