

Apports de la lexicographie réticulaire : le cas de la polysémie relationnelle

Alain POLGUÈRE
Université de Lorraine, CNRS, ATILF

Résumé

Les Systèmes Lexicaux sont un type particulier de réseaux lexicaux, dont la construction lexicographique repose sur les principes théoriques et descriptifs de la Lexicologie Explicative et Combinatoire. Dans cet article, nous expliquons pourquoi la lexicographie « réticulaire » des Systèmes Lexicaux – ou autres modèles analogues – favorise la mise au jour de phénomènes linguistiques fondamentaux que la lexicographie des dictionnaires tend à dissimuler derrière le rideau textuel de ses entrées lexicales. Nous utilisons pour cela le problème de la modélisation de la polysémie des vocables. Après avoir introduit notre problématique, nous faisons un rappel des caractéristiques formelles des Systèmes Lexicaux. Nous présentons ensuite la modélisation relationnelle de la polysémie des vocables telle qu’effectuée dans le Système Lexical appelé *Réseau Lexical du Français*. Finalement, nous examinons les phénomènes qu’une telle modélisation permet de traiter.

Mots clés : Lexicologie Explicative et Combinatoire, Système Lexical, *Réseau Lexical du Français* (RL-fr), lexicographie réticulaire, polysémie.

Abstract

Lexical Systems are a specific type of lexical networks whose lexicographic construction is based on theoretical and descriptive principles of Explanatory Combinatorial Lexicology. In this paper, we explain why the “reticular” lexicography of Lexical Systems – or other similar models – helps in revealing fundamental linguistic phenomena that dictionary-based lexicography tends to hide behind the textual curtain of its lexical entries. For this, we use the modeling of polysemy as illustration. After a general presentation of the topic under scrutiny, we state the formal characteristics of Lexical Systems. Next, we present the relational modeling of vocables’ polysemy implemented in a given Lexical System: the *French Lexical Network*. To conclude, we examine linguistic phenomena which can be better handled through the recourse to such approach.

Keywords: Explanatory Combinatorial Lexicology, Lexical System, *French Lexical Network* (fr-LN), reticular lexicography, polysemy.

Resumen¹

Los Sistemas Léxicos son un tipo particular de redes léxicas cuya construcción lexicográfica reposa sobre los principios teóricos y descriptivos de la Lexicología Explicativa y Combinatoria. En este artículo, explicamos por qué la lexicografía “reticular” de los Sistemas Léxicos –o de otros modelos análogos– favorece la actualización de fenómenos lingüísticos fundamentales que la lexicografía de los diccionarios tiende a disimular tras el telón textual de sus entradas léxicas. Nos serviremos, para ello, del problema de la modelización de la polisemia de los vocablos. Tras haber introducido nuestra problemática, recordaremos las características formales de los Sistemas Léxicos. A continuación, presentaremos la modelización relacional de la polisemia de los vocablos tal como esta se efectúa en el Sistema Léxico denominado *Réseau Lexical du Français* (Red Léxica del Francés). Finalmente, examinaremos los fenómenos que dicha modelización permite tratar.

Palabras clave: Lexicología Explicativa y Combinatoria, Sistema Léxico, *Red Léxica del Francés* (RL-fr), lexicografía reticular, polisemia.

1. Introduction

La lexicographie textuelle – celle qui construit des modèles lexicaux de type dictionnaire – a pour cible un modèle du lexique organisé par imbrication et ordonnancement linéaire d’éléments. Formellement, ce modèle est un très long texte structuré en parties, sous-parties et composants : « chapitres » correspondant aux initiales alphabétiques ou l’équivalent, entrées, articles, zones d’articles... Le dictionnaire possède une mise en page qui explicite ce qui constitue son fondement structural linéaire – sa macrostructure et sa microstructure – et met en arrière-plan la nature intrinsèquement relationnelle du lexique – ce qui a été théorisé dans une perspective métalexigraphique sous la désignation de *médiostucture* (Gows et Prinsloo, 1988). Par contraste, la lexicographie que l’on peut qualifier de *reticulaire* – celle qui construit des réseaux d’unités lexicales – a pour cible un modèle organisé comme l’est un réseau social sur Internet ou, plus généralement, comme tout phénomène naturel relationnel formant un grand *graphe de terrain* (Gaume, 2004) : un vaste réseau non taxonomique constitué d’un ensemble d’unités lexicales connectées par des relations paradigmatiques et syntagmatiques. De ce fait, il n’est pas exagéré de dire que tout modèle réticulaire du lexique est en tout premier lieu une médiostucture.

¹ Merci à Xavier Blanco pour cette traduction en espagnol.

Dans cet article, nous nous focalisons sur un type particulier de réseaux lexicaux appelés *Systèmes Lexicaux*. Nous expliquons pourquoi, selon nous, la lexicographie des Systèmes Lexicaux – ou de tout autre modèle analogue – favorise la mise au jour de phénomènes linguistiques fondamentaux que la lexicographie des dictionnaires tend à dissimuler derrière un rideau textuel. Nous utilisons pour cela le problème spécifique du traitement de la polysémie des vocables au sein d’une modélisation réticulaire du lexique. Notre argumentation s’articule en trois parties : rappel des caractéristiques formelles des modèles réticulaires du lexique de type Systèmes Lexicaux et des particularités de la mise en œuvre de leur construction lexicographique (section 2) ; présentation de la modélisation relationnelle de la polysémie des vocables telle qu’effectuée dans le Système Lexical appelé *Réseau Lexical du Français* ou RL-fr (section 3) ; phénomènes qu’une telle modélisation permet de traiter et qui sont occultés dans l’approche traditionnelle des dictionnaires textuels (section 4). Pour faciliter la compréhension de ce texte, nous faisons maintenant un bref rappel de notre terminologie de base relativement aux *entités lexicales* et, ce qui en est indissociable, de nos conventions d’écriture².

Les *unités lexicales*, ou *lexies*, sont les unités de modélisation lexicographique. Elles se regroupent naturellement au sein de *vocables* (polysémiques), ce qui conduit à les distinguer par des numéros d’acception. Les noms de vocables et de lexies s’écrivent en petites majuscules. Ainsi, le vocable COUDE, qui nous servira pour illustrer notre propos dans la section 3, regroupe les quatre lexies suivantes :

- COUDE **I** [*Elle s’est cogné le coude.*] ;
- COUDE **II** [*Sa veste est tout usée aux coudes.*] ;
- COUDE **III.1** [*La rivière forme un large coude dans la vallée.*] ;
- COUDE **III.2** [*La canalisation se termine par un coude en plastique.*].

Deux lexies sont regroupées au sein d’un même vocable si elles s’expriment par les mêmes signifiants et si elles possèdent une intersection de sens significative. Notre approche étant synchronique, c’est l’existence

² La quasi-totalité de la terminologie et des conventions d’écriture dont nous faisons usage est introduite de façon systématique dans Polguère (2016a).

d'une relation sémantique dans l'état actuel la langue (où dans la tranche temporelle considérée) qui importe.

Les lexies sont soit des *lexèmes* – unités monolexicales comme COUDE **I**, COUDE **II**, etc. –, soit des *locutions* – unités polylexicales telles que **I** USINE À GAZ **I** [L'ancienne usine à gaz s'étendait sur une parcelle de 9 600 m².] et **I** USINE À GAZ **II** [Cette réforme gouvernementale est une véritable usine à gaz]. Comme on le voit, les noms de locutions sont encadrés par les symboles **I** ... **I**.

1. Modèles lexicographiques de type Systèmes Lexicaux

La lexicographie ne vise pas la seule construction de dictionnaires – ouvrages répertoriant les vocables de la langue et leur description. Sa finalité est, de façon plus générale, la construction de modèles du lexique des langues. Si traditionnellement ces derniers ont pris la forme de dictionnaires, les bouleversements technologiques qui se sont produits au XX^e siècle ont rendu possible la mise en œuvre d'une nouvelle lexicographie, libérée des contraintes très fortes imposées par la structure textuelle des dictionnaires. Il s'agit d'une lexicographie qui, d'une part, reconnaît comme primordiale la structure relationnelle du lexique et, d'autre part, se fixe pour objectif le *tissage de réseaux lexicaux* – par contraste à la *rédaction de dictionnaires*. Nous avons déjà présenté à plusieurs reprises les motivations et principes de cette lexicographie que l'on peut qualifier de *réticulaire* et nous ne reviendrons pas ici sur cette discussion – cf. notamment Polguère (2012, 2016b).

Il serait erroné de penser que la lexicographie réticulaire a émergé d'une soudaine perception de la nature relationnelle du lexique, rendue possible par le potentiel descriptif des modèles informatisés. Une large proportion des théoriciens et praticiens de la lexicographie des dictionnaires a bien perçu la contradiction structurale qui existe entre l'organisation des dictionnaires et celle du lexique. C'est pourquoi de nombreuses expérimentations et innovations ont eu lieu par le passé ; pour ne citer que deux d'entre elles : le *Longman Language Activator* (Summers, 1993), pour l'anglais, et le *Dictionnaire du français usuel* (Picoche et Rolland, 2002), pour le français. Toutes deux ont cherché à faire éclater la barrière de l'organisation linéaire de l'information dictionnaire en proposant des structures qui conditionnent une navigation multidimensionnelle dans le réseau lexical de la langue.

Il ne suffit pas de postuler que les modèles lexicaux doivent être structurés en réseau pour pouvoir mettre en œuvre une véritable lexicographie réticulaire. Seules une théorisation et formalisation de la structure relationnelle du lexique, reposant sur une théorie linguistique générale, permet l'implémentation rigoureuse d'une telle lexicographie. Par *théorie linguistique générale*, nous entendons une théorie qui prend en charge, notamment, la modélisation (i) du lexique, (ii) de la grammaire et (iii) de l'interdépendance des deux. Les phénomènes à modéliser relèvent tout aussi bien des modules sémantique, syntaxique, morphologique et phonologique de la langue. Finalement, la prise en compte de la *phraséologie* de la langue doit être au cœur de la théorisation et, par voie de conséquence, de la modélisation du lexique. Les *Systèmes Lexicaux* (Polguère, 2009, 2014) sont un type particulier de réseaux lexicaux, dont la construction lexicographique repose sur les principes théoriques et descriptifs de la *Lexicologie Explicative et Combinatoire* (Mel'čuk et coll., 1995)³ et, plus généralement, de l'approche linguistique *Sens-Texte* (Mel'čuk, 2016).

Polguère (2014) explicite les grands principes d'une lexicographie orientée vers la construction de tels modèles. Nous nous contentons ici d'énumérer les quatre propriétés formelles fondamentales des Systèmes Lexicaux, telles que détaillées dans Polguère (2014, 6–9).

- 1) Un Système Lexical d'une langue L – désormais, SL_L – est un graphe orienté dont les nœuds sont, prioritairement, les unités lexicales de L (lexèmes ou locutions) et les arcs correspondent, prioritairement, à deux familles de relations lexicales :
 - a) relations paradigmatisques de *dérivations sémantiques* – de la source à la cible de chaque dérivation (*larmes* → *pleurer*, *larmes* → *larmoyant...*) ;
 - b) relations syntagmatiques du type base → collocatif au sein de *collocations*, au sens de Hausmann (1979)⁴.

³ En plus des références bibliographiques citées ici, on pourra consulter le site Web des Systèmes Lexicaux à l'adresse suivante : <https://lexical-systems.atilf.fr> (consulté 30/05/2020).

⁴ La flèche représente ici la dépendance fonctionnelle du collocatif vis-à-vis de la base de la collocation : il est sélectionné en fonction de la base. D'un point de vue syntaxique, un collocatif peut dépendre syntaxiquement de la base (*larmes de crocodile*) ou la gouverner (*verser des larmes*).

Ces relations sont modélisées au moyen des *fonctions lexicales* Sens-Texte (Mel'čuk, 1996 ; Mel'čuk et Polguère, à paraître), qui tissent donc l'ossature relationnelle globale de SL_L .

- 2) Les nœuds de SL_L – qui modélisent les lexies de L et autres entités lexicales⁵ – sont des objets formels complexes. Ils encapsulent un contenu informationnel riche correspondant au contenu d'un article lexicographique (caractéristiques grammaticales, définition, exemples lexicographiques, etc.).
- 3) SL_L est un *modèle non taxonomique* du lexique de L , en ce sens qu'il n'organise pas ce dernier en une hiérarchie de classes et sous-classes de lexies, contrairement à WordNet (Miller et coll., 1990) ou aux modèles qu'il a inspirés. Formellement, il s'agit d'un grand *graphe de terrain* (Gaume, 2004), dont la structure mathématique a été conceptualisée sous le nom de *graphe petit-monde* (Watts et Strogatz, 1998). Une telle structure est analogue à celle des réseaux sociaux, où une constellation apparemment anarchique d'entités individuelles s'organise en un système complexe d'agrégats « naturels » (*clusters*) selon une répartition topologique caractéristique. Les travaux de B. Gaume, notamment, ont démontré que l'analyse topologique de structures lexicales de ce type permettait de modéliser informatiquement des phénomènes linguistiques fondamentaux, dont la proximité sémantique entre entités lexicales – cf. la notion de *proxémie* (Gaume, 2004 ; Gaume et coll., 2008).

Pour rendre plus explicite la nature formelle des Systèmes Lexicaux, nous donnons dans la Figure 1 une visualisation d'un extrait du *Réseau Lexical du Français*, un Système Lexical dont il sera question dans la section suivante. Cette visualisation repose sur une géométrisation de l'environnement topologique des deux acceptions du vocable « USINE A GAZ », dont la polysémie a été mentionnée à la fin de la section 1. Elle contient également l'affichage du contenu informationnel d'un des nœuds lexicaux – l'acception métaphorique

⁵ Les entités lexicales autres que les lexies peuvent être, par exemple, des clichés linguistiques (Mel'čuk, 2015 ; Polguère, 2016c), tels que *Après vous, Comment ça va ?*, etc.

「USINE À GAZ」⁶ du vocable – sous la forme (textuelle) d’un article lexicographique⁶.

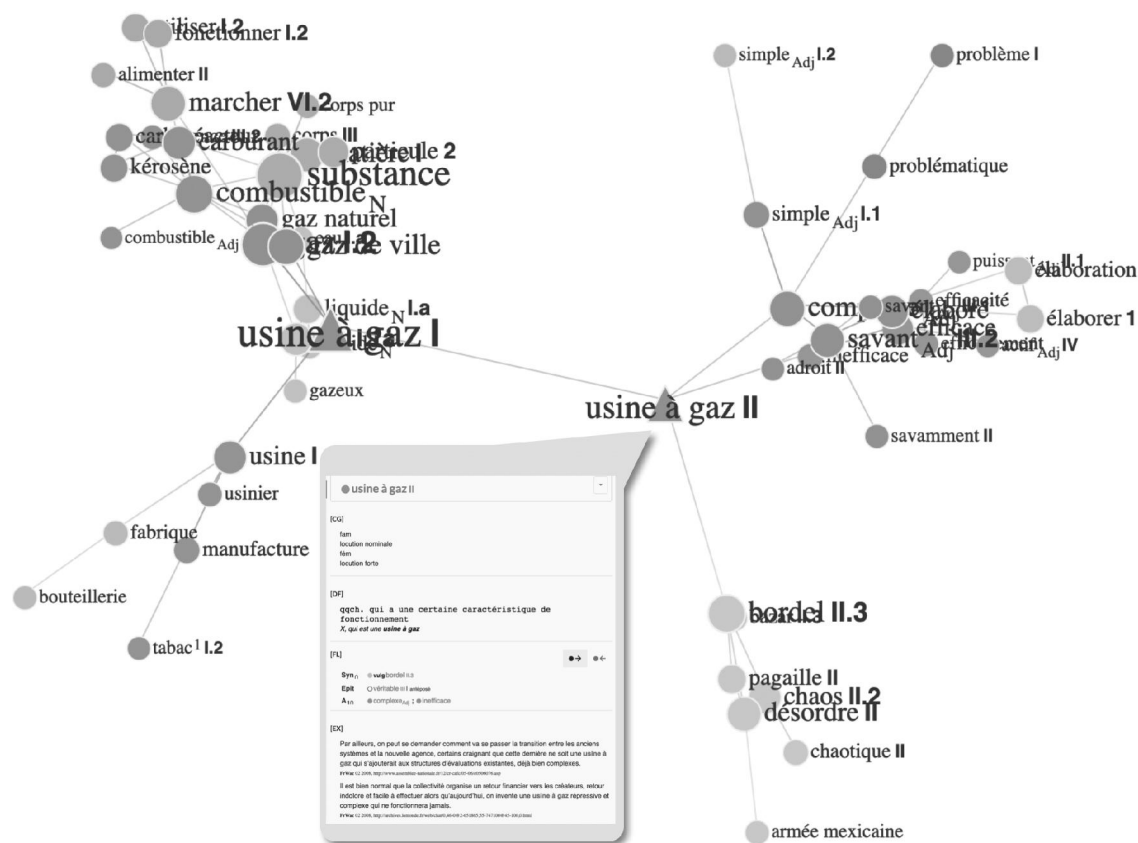


Figure 1. Structure d'un Système Lexical : autour du vocable 「USINE À GAZ」

4) La quatrième propriété fondamentale de SL_L est d'offrir une *modélisation relativiste* du lexique de la langue L , où il est possible de nuancer les différents éléments d'information encodés en leur associant un *degré de confiance* non maximal (c'est-à-dire, une probabilité de validité < 1.00).

Contrairement à l'approche standard de la Lexicographie Explicative et Combinatoire, la lexicographie des Systèmes Lexicaux repose sur le postulat que la modélisation du lexique ne doit pas se faire uniquement en termes d'indications présentées comme « vraies » dans l'absolu, cela pour au moins deux raisons. Tout d'abord, la confrontation avec les faits de

⁶ L'image de la Figure 1 est de taille trop réduite pour en permettre une véritable lecture. Elle sert avant tout à illustrer la structure informationnelle d'un Système Lexical. Les visualisations de graphes lexicaux présentées ici sont obtenues à partir de l'interface de navigation Spiderlex (Ollinger et coll., 2020). Pour une présentation de ce dernier et un accès au navigateur, voir <https://lexical-systems.atilf.fr/spiderlex/> (consulté 09/05/2020).

Parole peut pousser à des conclusions contradictoires sur l'état de la Langue qu'il est parfois nécessaire de refléter dans les modèles. Ensuite, et de façon concomitante, il est dans la nature de la Langue, et donc du lexique, d'être dans un état d'instabilité permanent – cf. la notion saussurienne de *mutabilité du signe linguistique* (Saussure, 1995 [1916], Première partie, Chap. II, § 2). Il est donc nécessaire d'avoir les moyens formels de relativiser l'applicabilité d'une information lexicale ou, plus généralement, d'une règle linguistique. Une bonne illustration de cela, sur le plan lexical, est la variation appelée *séparation phraséologique* dans Pausé et Polguère (2020), par laquelle une locution se mue en une collocation : par exemple, 「BOIRE LA TASSE」 → *avalier / boire / se prendre... une tasse*⁷. Dans un état donné de la langue, une locution peut être en train de subir une telle mutation et son existence même entre en concurrence avec celle de la collocation émergente. Il faut alors refléter cet état de fait en abaissant le degré de confiance associé à la déclaration de chacune des deux entités lexicales en compétition (locution *vs* collocation).

La lexicographie des Systèmes Lexicaux repose sur le postulat sous-jacent que les modèles qu'elle construit possèdent une plus grande « adéquation psycholinguistique » avec le *lexique mental* que les dictionnaires (Polguère, 2016b). Un tel postulat est bien entendu, par nature, impossible à démontrer. On ne peut que le valider par les exploitations (par exemple, pédagogiques) qui pourront être faites des Systèmes Lexicaux. Afin d'illustrer en quoi l'approche des Systèmes Lexicaux et, plus généralement, la lexicographie réticulaire, offre des moyens inédits de traiter la connaissance lexicale, nous nous focalisons, dans la suite de l'article, sur le cas du traitement de la polysémie dans un modèle réticulaire du type Système Lexical.

3. Intégration de la polysémie dans les Systèmes Lexicaux

3.1 Postulats sur la polysémie

La polysémie lexicale est l'objet de multiples débats et a donné lieu à un flot de publications, que ce soit pour relativiser la notion (Kilgarriff,

⁷ Plus que la croissance du paradigme des verbes gouverneurs de *tasse*, c'est le fait que ce nom commence à s'employer de façon isolée pour dénoter le fait en question qui démontre que la locution se mue en une collocation, dont la base est *tasse* et où *boire* fonctionne comme verbe support.

1997 ; Fodor, 1998 ; Recanati, 2004) ou pour en étayer les fondements (Apresjan, 1974 ; Geeraerts, 1990, 1993 ; Cruse, 1995 ; Pustejovsky, 1995 ; Goddard, 2000). Il n'est pas dans notre propos d'entamer ici une nouvelle discussion théorique et nous admettons les deux postulats suivants :

1. la polysémie des vocables est un fait linguistique établi ;
2. elle peut et doit être systématiquement modélisée dans le cadre de la construction de modèles lexicographiques.

Pour les personnes sérieusement impliquées dans la tâche de modélisation du lexique et dans ses applications, ces postulats peuvent sembler des évidences. Il ne faut cependant pas perdre de vue que les courants de pensée dominants sont fréquemment promulgués par ceux qui disent plutôt que par ceux qui font, et qu'une conception malheureusement très présente de la polysémie en linguistique et sciences cognitives est assez bien illustrée par la citation suivante⁸ :

My theory is that there is no such thing as polysemy. The appearance that there is a problem is generated by the assumption that there are definitions; if you take the assumption away, the problem disappears. (Fodor, 1998, 53)

Enfin, il est important de rappeler que notre perspective sur la polysémie est de nature synchronique. Nous cherchons à nous détacher le plus possible de considérations diachroniques, qui relèvent d'une autre notion de polysémie, même si nous sommes convaincu que la modélisation de la polysémie dans les Systèmes Lexicaux peut permettre de mieux modéliser à leur tour les phénomènes dynamiques à l'œuvre dans la structuration du lexique⁹.

Dans un premier temps (3.2), nous revisitons la notion de polysémie pour l'approcher dans une perspective relationnelle,

⁸ Il existe un courant de pensée assez populaire en linguistique théorique (beaucoup moins présent en linguistique descriptive) qui nie l'existence même de la polysémie, dans le contexte d'une approche de la signification lexicale que l'on peut qualifier de *contextualisme lexical*. Pour une tentative d'explication des motifs d'une telle remise en question de la polysémie, on pourra se reporter à Polguère (2015).

⁹ La perspective synchronique sur la polysémie peut, dans une très large mesure, être méthodologiquement autonome de la perspective diachronique. Il est cependant clair qu'une compréhension entière des mécanismes linguistiques mis en jeu dans la polysémie au sens large nécessite que les deux approches soient ultimement mises en relation.

compatible avec l'approche des Systèmes Lexicaux. Ensuite (3.3), nous expliquons comment la polysémie est modélisée dans le Système Lexical lexicographiquement construit appelé *Réseau Lexical du Français*¹⁰.

3.2 Polysémie en tant que phénomène relationnel

Le phénomène polysémique est traditionnellement conceptualisé en tant que propriété de mots ou plus précisément, dans notre terminologie, de vocables. On parle de la *polysémie du vocable V*, qui est la propriété de V d'être un ensemble constitué de plus d'une unité lexicale. La polysémie d'un vocable est ainsi le pendant dans la Langue, au sens Saussurien, de l'ambiguïté d'une occurrence de signifiant dans la Parole. Cela fait que la polysémie est très fréquemment définie à partir de la notion d'ambiguïté, comme le soulignent Paulin et Béjoint (2008, 8) :

Polyseny is one of the major problems encountered by semanticists, who sometimes prefer to refer to it as "ambiguity".

Le fondement logique de la polysémie n'est cependant ni l'ambiguïté (Picoche, 1989), ni une **propriété** d'un vocable, mais une **relation** entre lexies, relation que nous appelons *copolysémie*¹¹. Elle peut se définir ainsi¹² :

La copolysémie entre la lexie L_1 et la lexie L_2 – symbolisée par $L_1 \rightarrow L_2$ – est la relation formelle et sémantique¹³ qui lie L_2 à L_1 et qui conditionne leur regroupement au sein d'un même vocable.

Pour illustrer notre propos, considérons la polysémie du vocable COUDE, décrite de façon énumérative dans la section 1. Cette polysémie repose sur un système de trois relations de copolysémie connectant quatre *copolysèmes* présenté dans la Figure 2.

¹⁰ Les sections 3.2 et 3.3 reprennent certains éléments présentés, en anglais, dans Polguère (2018). Faute de place, nous ne pouvons décrire ici au complet la modélisation de la polysémie intégrée dans le *Réseau Lexical du Français*. Pour cela, on se reportera à Polguère (2018).

¹¹ Les études de linguistique, lorsqu'elles veulent référer à la relation sous-jacente à la polysémie des vocables, parlent de *relation de polysémie*. Il nous semble préférable de disposer de deux termes établissant une claire distinction entre une relation lexicale (la copolysémie) et une propriété résultante des vocables (la polysémie).

¹² Traduction française de la définition formulée en anglais dans Polguère (2018, 794).

¹³ Il s'agit de la relation d'identité de signifiants et d'intersection significative de sens que nous avons mentionnée à la fin de la section 1, lors de la caractérisation de la notion de vocable.

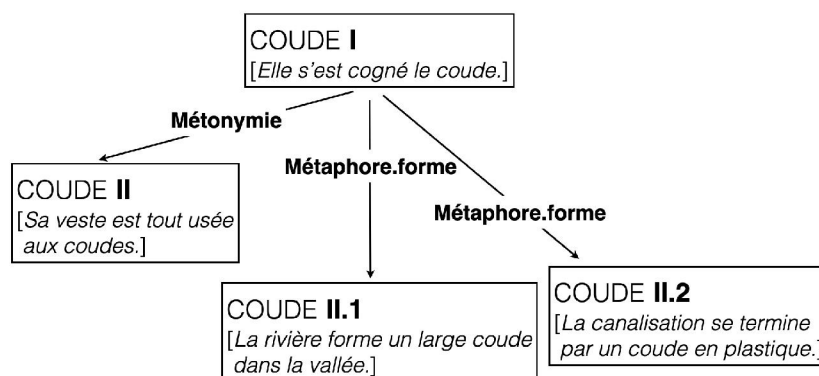


Figure 2. Structure polysémique du vocable COUDE

À notre connaissance, aucun modèle lexicographique existant – à l’exception du *Réseau Lexical du Français*, dont il sera question en 3.3 – ne traite systématiquement la structure relationnelle des vocables polysémiques en explicitant les relations de copolysémie. Les dictionnaires de langue, notamment, se contentent de préciser de façon ponctuelle la nature métonymique, métaphorique, etc., de certaines acceptions, quand ils ne passent pas de telles relations entièrement sous silence¹⁴. Même un modèle très formalisé comme le WordNet de Princeton, pour l’anglais américain, ne fait qu’énumérer et numéroter les acceptions, sans les mettre en relation et sans perspective sur la structuration polysémique des vocables¹⁵. C’est principalement dans le cadre de travaux théoriques que l’on voit apparaître des graphes de structures polysémiques. Il peut s’agir d’études sur la diachronie, c’est-à-dire sur la dynamique de la polysémie – cf. les graphes dynamiques proposés par François (2020) – ou d’études sur la structuration formelle du lexique – cf. le rôle de la polysémie dans la structuration en graphe petit-monde du lexique dans Sigman et Cecchi (2002).

¹⁴ Le *Trésor de la Langue Française* (TLFi, 2012) est une exception en lexicographie française. Il s’agit probablement du dictionnaire de référence qui tend à fournir le plus d’informations sur la nature de la dérivation polysémique des acceptions dans ses entrées, bien entendu dans une perspective diachronique. On pourra par exemple se reporter à l’entrée COUDE dans ce dictionnaire et la comparer aux mêmes entrées dans d’autres dictionnaires de langue.

¹⁵ Cf. la documentation de WordNet en ligne : « *WordNet senses are ordered using sparse data from semantically tagged text. The order of the senses is given simply so that some of the most common uses are listed above others (and those for which there is no data are randomly ordered). The sense numbers and ordering of senses in WordNet should be considered random for research purposes.* », <https://wordnet.princeton.edu/frequently-asked-questions> (consulté 02/05/2020).

3.3 Polysémie des vocables dans le *Réseau Lexical du Français (RL-fr)*

Le *Réseau Lexical du Français* – désormais *RL-fr* – est un modèle du lexique français construit lexicographiquement selon l’approche des Systèmes Lexicaux (section 2). Même si la construction du *RL-fr* est un projet à long terme, il n’est aucunement dans son état actuel un échantillon lexicographique. Les statistiques sur sa couverture en tant que graphe lexical, au moment où nous écrivons ces lignes, sont résumées dans le Tableau 1.

28.730 nœuds lexicaux – principalement des lexies (lexèmes et locutions) – regroupés en 18.354 vocables polysémiques
81.174 arcs de relation lexicale – en majorité (60.910) des relations de fonctions lexicales paradigmatiques et syntagmatiques, mais avec une proportion significative de liens de copolysémie (9.141)
32.642 citations extraites de corpus servant à illustrer les lexies décrites et formant un corpus de citations de plus de 1,2 million de mots-occurrences

Tableau 1. Statistiques sur le graphe lexical du *RL-fr*

Les deux premières lignes du Tableau 1 donnent un aperçu de la couverture de *RL-fr* en tant que graphe lexical (ensemble de nœuds et d’arcs connectant ces nœuds). La troisième ligne du tableau sert à donner un aperçu de la richesse de l’information lexicographique qui peut être associée aux nœuds lexicaux : ici la connexion de chaque nœud du graphe à des citations extraites de corpus et qui servent d’exemples lexicographiques. Les données du *RL-fr* sont accessibles de trois façons¹⁶ :

- navigation graphique en ligne dans les données actuelles du *RL-fr*, mises à jour quotidiennement, au moyen du navigateur Spiderlex (*cf.* note 6) ;
- export des données lexicographiques en accès libre sur la plateforme ORTOLANG (Ollinger et Polguère, 2020) ;

¹⁶ Pour plus d’informations, on peut se reporter aux pages « Navigateur Spiderlex » et « Téléchargements » du site des Systèmes Lexicaux, déjà mentionné dans la note 3.

- export des citations lexicographiques sous forme de ressource autonome (BEL-RL-fr) en accès libre sur la plateforme ORTOLANG¹⁷.

Même si les arcs d'un Système Lexical correspondent prioritairement à des liens de fonctions lexicales (*cf.* section 2), les statistiques du Tableau 2 indiquent qu'une part non triviale des arcs du RL-fr – plus de 11 % – correspondent à des liens de copolysémie tissés entre les acceptions des vocables décrits. Le tissage de ces liens permet d'intégrer dans le Système Lexical qu'est le RL-fr des micrographes correspondant à des structures polysémiques de vocables, comme celle de la Figure 2 (section 3.2), pour COUDE. Cette structure est relativement simple et « plate ». De nombreux vocables sont cependant organisés de façon plus élaborée, même dans le cas de vocables à la polysémie moyennement riche, comme JAMBE, Figure 3, pour rester dans le champ lexical des éléments du corps.

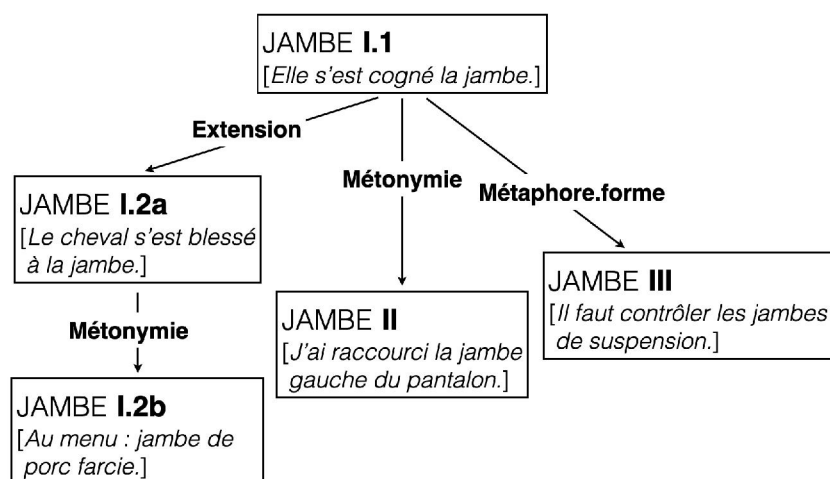


Figure 3. Structure polysémique du vocable JAMBE

Les relations de copolysémie qui étiquettent les arcs des structures polysémiques de vocables – *cf.* Figures 2 et 3 – constituent un modèle de la polysémie construit de façon inductive par les lexicographes du RL-fr. Même s'il trouve essentiellement son origine dans un travail sur le français, ce modèle a été élaboré dans une perspective multilingue. Il est d'ailleurs appliqué simultanément à deux autres Systèmes Lexicaux : le *Réseau Lexical de l'Anglais* (Gader et coll., 2014) et le *Réseau Lexical du Russe* (Krylosova, 2017). Faute de place, nous ne pouvons présenter et justifier

¹⁷ Documentation en ligne et téléchargement de la base : <https://www.ortolang.fr/market/corpora/examples-ls-fr/v1> (consulté 23/03/2020).

en détail le système de relations de copolysémie qui a été identifié. Il est synthétisé dans le Tableau 2 et fait l'objet d'une discussion approfondie dans Polguère (2018).

Poids	Type	Sous-type	Dérivation
Fort	Conversion		Oui
Fort	Causation		Oui
Fort	Résultat		Oui
Fort	Spécialisation	Sous-sens	Oui
Fort	Généralisation		Oui
Fort	Métonymie	Ensemble de, Partie de	Oui
Faible	Extension		Oui
Faible	Antiphrase		Oui
Faible	Intersection		Non
Non cohésif	Métaphore	Comme si, Comme si méton, Comportement, Couleur, Fonction, Fonctionnement, Forme, Mouvement, Position, Structure, Utilisation	Oui
Non cohésif	Grammaticalisation		Non
Non cohésif	Codérivation		Non

Tableau 2. Relations de copolysémie dans le RL-fr¹⁸

¹⁸ Ce tableau est adapté de Polguère (2018, 804) et intègre quelques changements effectués dans le modèle de la polysémie du RL-fr depuis l'écriture de l'article (nouveaux sous-types Couleur et

Chaque ligne du Tableau 2 correspond à une relation de copolysémie donnée ; les colonnes contiennent la caractérisation des relations. Expliquons brièvement le contenu de chacune de ces colonnes.

La colonne *Poids* conditionne l'ordre d'énumération des types de relations de copolysémie $L_1 \rightarrow L_2$. Chaque type de relation $L_1 \rightarrow L_2$ se voit associé un poids sémantique, qui reflète l'importance de l'intersection de sens entre la lexie source L_1 et la lexie cible L_2 de la relation. Le poids est soit *Fort*, soit *Faible*, soit *Non cohésif* – c'est-à-dire qu'il ne participe pas à la cohésion sémantique au sein du Système Lexical.

La colonne *Type* contient les noms des familles de relations $L_1 \rightarrow L_2$. Dans la mesure du possible, ces noms sont empruntés à une terminologie couramment employée en lexicologie et lexicographie. Contrairement à la pratique courante, cependant, les termes utilisés dans le modèle du Tableau 2 sont explicitement caractérisés, en tout premier lieu, en fonction de la nature de la relation de sens qu'entretiennent L_1 et L_2 (Polguère, 2018, 805-816). La *définition lexicographique* joue ici un rôle central. Notre approche étant synchronique, le lexicographe postule une relation donnée $L_1 \rightarrow L_2$ à partir (i) des deux définitions qu'il élabore pour L_1 et L_2 et (ii) de l'évaluation qu'il fait du rapport entre ces deux définitions. Il n'est pas nécessaire que le lexicographe mette véritablement noir sur blanc les définitions en question. Nous considérons qu'une compétence que doit posséder tout lexicographe est celle de pouvoir ébaucher « à la volée » une définition lexicographique, en appliquant de façon quasi instantanée des principes analytiques et descriptifs rigoureux, tels que ceux décrits dans Mel'čuk et Polguère (2018). La capacité d'ébaucher mentalement une définition lexicographique logiquement structurée est, dans le domaine lexical, aussi importante que celle de pouvoir associer de façon quasi immédiate un arbre syntaxique à une phrase, dans le domaine de la syntaxe. Elle n'est d'ailleurs pas utile qu'aux lexicographes : elle trouve son application notamment en enseignement de la langue maternelle ou seconde, et dans tout contexte où il nous est demandé d'improviser l'explication du sens d'une lexie.

Mouvement). Le modèle de la polysémie présenté dans ce tableau est incomplet et évolue graduellement, au fur et à mesure de la progression du travail lexicographique.

La colonne *Sous-Type*, comme son nom l'indique, apporte une spécification additionnelle de certaines relations de copolysémie. Dans l'état actuel du modèle, seuls trois types de relations se voient associer des sous-types, et ceux-ci ne sont obligatoirement spécifiés (*cf.* le gras dans le Tableau 2) que pour la relation de *Métaphore*. La liste des sous-types est encore très incomplète. Néanmoins, il nous semble souhaitable de limiter au maximum la granularité du système pour lui garder sa généralité et son applicabilité dans un contexte interlinguistique.

Finalement, la colonne *Dérivation* indique si le type de relation considéré est oui ou non une dérivation véritable $L_1 \rightarrow L_2$. Il existe en effet certains cas où le lien de copolysémie ne peut être caractérisé comme un cas de « production sémantique » de L_2 à partir de L_1 – pour plus de détails, se reporter à Polguère (2018).

La décision de tisser les liens de copolysémie au sein des Systèmes Lexicaux est une initiative tardive dans le projet lexicographique du RL-fr, qui était focalisé au départ sur le tissage des relations de fonctions lexicales paradigmatiques et syntagmatiques. Il est graduellement apparu que l'ensemble des structures polysémiques des vocables était véritablement *the elephant in the room* au sein de notre modèle relationnel. Le tissage de la copolysémie a ainsi émergé comme une évidence du fait du rôle fondamental joué par la polysémie dans la structuration du lexique. Ce point fait l'objet de la prochaine section, qui nous sert en même temps de conclusion.

4. Apports de la modélisation réticulaire de la polysémie

4.1 Apports descriptifs

Du point de vue descriptif ou, plus précisément, lexicographique, la modélisation réticulaire de la polysémie permet de contourner les limitations de l'approche énumérative des dictionnaires. Dans ceux-ci, trois outils sont utilisés pour tenter de refléter la structuration polysémique des vocables :

1. ordonnancement des acceptions dans leur énumération selon une logique qui peut être chronologique, sémantique, « psychologique » (*cf.* la notion de *dominant reading*), statistique ou un panachage de ces perspectives ;

2. système de nommage des acceptions qui tente de rendre compte à la fois de la hiérarchisation de celles-ci et de leur proximité relative – chiffres romains, chiffres arabes, lettres, puces, losanges, etc. ;
3. mention explicite de la nature de la relation qui unit une acception à une autre, la source de la relation n'étant généralement pas identifiée explicitement – par exemple, *métaph.* ou *méton.*

La polysémie de JAMBE décrite par un graphe dans la Figure 3 (section 3.3) peut ainsi être reformulée dans le Tableau 3 de façon énumérative, sur le mode dictionnaire, en faisant usage des trois outils identifiés ci-dessus.

JAMBE **I.1** [*Elle s'est cogné la jambe.*]

JAMBE **I.2a** ext. de **I.1** [*Le cheval s'est blessé à la jambe.*]

JAMBE **I.2b** méton. de **I.2a** [*Au menu : jambe de porc farcie.*]

JAMBE **II** méton. de **I.1** [*J'ai raccourci la jambe gauche du pantalon.*]

JAMBE **III** métaph. de **I.1** [*Il faut contrôler les jambes de suspension.*]

Tableau 3. Énumération de la polysémie de JAMBE

Outre le fait que *ext.*, *méton.*, *métaph.*, etc., ne sont pratiquement jamais suivis dans les dictionnaires par l'identification de la source du lien de copolysémie, la description ci-dessus ne donne aucun accès à la structure même du vocable, puisqu'elle n'est pas fondée sur une formalisation relationnelle. Les éléments de nature textuelle du Tableau 3 **reflètent**, plutôt qu'ils ne **modélisent**, la structure du vocable. Seule une lexicographie réticulaire, telle que celle que nous avons explicitée plus haut, réifie les structures polysémiques, qui peuvent alors être automatiquement visualisées (sous la forme des Figures 2 et 3), analysées, triées, révisées, etc.

Un autre apport important de la lexicographie réticulaire au traitement de la polysémie concerne le problème de la distinction des acceptions et des critères permettant de l'établir (Cruse, 1995 ; Mel'čuk, 2013, Chapitre 11, section 3.2). Sans revenir ici sur les critères traditionnellement proposés (cooccurrence compatible, etc.), mentionnons qu'ils se basent sur des faits de langue qui trouvent en

partie leur source dans le *profil relationnel* des lexies exprimant chaque sens : relations paradigmatiques et syntagmatiques privilégiées que ces lexies entretiennent avec d'autres. Nous postulons que le critère primordial pour déterminer si deux contenus lexicaux correspondent à deux lexies distinctes copolysèmes est l'identification de deux *agrégats lexicaux* (angl. *lexical clusters*) distincts associés à chacun des contenus lexicaux en question au sein du Système Lexical de la langue (Polguère, 2018, 803). Cette approche a été appliquée systématiquement dans la lexicographie du RL-fr. Elle repose sur le postulat que ce qui justifie qu'une lexie est bien une unité lexicale dans une langue donnée, c'est avant tout le fait que cette langue singularise le sens correspondant en le positionnant dans un micro-système de relations lexicales qui établissent son profil dans la structure du lexique. Bien entendu, ce critère d'agrégat lexical ne suffit pas à établir la signature complète d'une lexie dans le lexique ; il faut aussi tenir compte des spécificités morphologiques, formelles, etc., qui peuvent caractériser chacune des acceptions d'un vocable. Nous pensons cependant que le fait d'être associé avec un agrégat lexical caractéristique – de posséder un profil d'« amis » distinct dans le réseau social des lexies de la langue – est la caractéristique première qui singularise chaque lexie, et donc chaque acception d'un vocable.

4.2 Apports théoriques

Sur le plan théorique, le fait de réifier la structure polysémique des vocables dans une approche réticulaire permet de ne plus considérer les liens de copolysémie uniquement en isolation, mais aussi en fonction de leur position dans l'économie générale de la polysémie de chaque vocable. Prenons comme illustration le cas de l'étude du phénomène fondamental qu'est la *polysémie régulière* (Aprésjan, 1974 ; Barque, 2008). Si l'on dispose de modèles lexicographiques réticulaires du type RL-fr, cette étude peut se mener non sur des familles de relations individuelles $L_1 \rightarrow L_2$, mais sur des *motifs de graphes* (Ollinger, 2014) de copolysémie. Cela est important, car les paramètres qui rendent possible une dérivation copolysémique donnée $L_1 \rightarrow L_2$ ne sont pas uniquement dépendants du rapport sémantique qu'entretiennent L_1 et L_2 ; ils dépendent aussi du contexte global de la polysémie du vocable au sein duquel se manifeste $L_1 \rightarrow L_2$. Nous ne sommes pas en mesure, faute de

recherche menée sur la question, de démontrer ce que nous avançons ici. Il s'agit cependant d'une perception qui s'est graduellement développée chez tous les lexicographes du RL-fr ayant effectué un travail soutenu de tissage des liens de copolysémie : une relation de copolysémie donnée ne prend complètement son sens et sa valeur que dans la structure globale du vocable où elle se manifeste. Ce fait nous semble devoir être pris en compte, tout particulièrement, si l'on veut élargir la perspective lexicographique au-delà de la seule modélisation structurale du lexique, pour l'amarrer à la modélisation des processus cognitifs mis en œuvre dans l'acquisition et l'utilisation de la connaissance lexicale.

Pour conclure, mentionnons un apport théorique purement formel du tissage des liens de copolysémie. Ces derniers participent de façon significative à la cohésion du Système Lexical en tant que graphe petit-monde (section 2). Pour une démonstration de ce fait, on pourra consulter l'expérimentation menée à partir de WordNet par Sigman et Cecchi (2002). Il est justifié de dire qu'un graphe lexical où sont tissés les liens de copolysémie possède les caractéristiques formelles d'un graphe petit-monde à un plus haut degré qu'un graphe qui n'inclut pas cette information. Il est donc plus compatible avec les calculs topologiques du type de ceux proposés par Gaume (2004), calculs qui ont démontré leur efficacité dans le traitement des grands graphes lexicaux. Il ne s'agit pas d'un fait anecdotique. Dans le domaine lexical, la validité d'un modèle formel s'évalue tout particulièrement en fonction de sa calculabilité. Le lexique est un système relationnel immense, très riche, au sein duquel le Locuteur parvient cependant à naviguer avec une extrême sûreté et rapidité lorsqu'il construit ses énoncés. La lexicographie réticulaire vise la construction de modèles lexicaux qui sont une hypothèse théorique sur la « géométrie du lexique » qui rend possible une telle navigation (Polguère, 2016b).

Références bibliographiques

APRESJAN, J. D., Regular Polysemy, *LINGUISTICS*, 1974, **142**, 5-32.

BARQUE, L., *Description et formalisation de la polysémie régulière du français*, thèse de doctorat, Paris, Université Paris 7, 2008.

CRUSE, D. A., Polysemy and Related Phenomena from a Cognitive Linguistic Viewpoint, in SAINT-DIZIER, P., VIEGAS, E., *Computational Lexical Semantics*, Cambridge *et al.*, Cambridge University Press, 1995, 33-49.

- FODOR, J. A., *Concepts. Where Cognitive Science Went Wrong*, Oxford Cognitive Science Series, Oxford, Oxford University Press, 1998.
- FRANCOIS, J., Pour un retraitement informatisé et dynamique des notices historiques du TLF, *CAHIERS DE LEXICOLOGIE*, 2020, 117, 55-92.
- GADER, N., OLLINGER, S., POLGUÈRE, A., One Lexicon, Two Structures : So What Gives ?, ORAV, H., FELLBAUM, C. D., VOSSEN, P., *Proceedings of the Seventh Global Wordnet Conference (GWC2014)*, Tartu, Global WordNet Association, 2014, 163-171.
- GAUME, B., Balades aléatoires dans les Petits Mondes Lexicaux, *I3 INFORMATION INTERACTION INTELLIGENCE*, 2004, **4(2)**, 39-96.
- GAUME, B., DUVIGNAU, K., PRÉVOT, L., DESALLE, Y., Toward a cognitive organization for electronic dictionaries, the case for semantic proxemy, *Proceedings of the workshop on Cognitive Aspects of the Lexicon (COGALEX 2008)*, 2008, 86-93.
- GODDARD, C., Polysemy. A Problem of Definition, in RAVIN, Y., LEACOCK, C., *Polysemy: Theoretical and Computational Approaches*, Oxford, Oxford University Press, 2000, 129-151.
- GEERAERTS, D., The lexicographical treatment of prototypical polysemy, in TSOHATZIDIS, S. L., *Meanings and Prototypes (RLE Linguistics B: Grammar). Studies in Linguistic Categorization*, Londres, Routledge, 1990, 195-210.
- GEERAERTS, D., Vagueness's puzzles, polysemy's vagaries, *COGNITIVE LINGUISTICS*, 1993, **4(3)**, 223-272.
- GOUWS, R. H., PRINSLOO, D. J., Cross-Referencing as a Lexicographic Device, *LEXIKOS*, 1988, **8**, 17-36.
- HAUSMANN, F. J., Un dictionnaire des collocations est-il possible ?, *TRAVAUX DE LITTÉRATURE ET DE LINGUISTIQUE DE L'UNIVERSITÉ DE STRASBOURG*, 1979, **XVII(1)**, 187-195.
- KILGARRIFF, A., I don't believe in word senses, *COMPUTERS AND THE HUMANITIES*, 1997, **31**, 91-113.
- KRYLOSOVA, S., Du projet d'élaboration d'un Réseau Lexical du Russe (RL-ru), *KARPOVSKIE NAUČNYE ČTENIJA*, 2017, **11(1)**, 243-246.
- MEL'ČUK, I., Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon, in WANNER, L., *Lexical Functions in Lexicography and Natural Language Processing*, Language Companion Series 31, Amsterdam / Philadelphie, John Benjamins, 1996, 37-102.
- MEL'ČUK, I., *Semantics: From Meaning to Text. Volume 2*, Studies in Language Companion Series 135, Amsterdam/Philadelphie, John Benjamins, 2013.
- MEL'ČUK, I., Clichés, an Understudied Subclass of Phrasemes, *YEARBOOK OF PHRASEOLOGY*, 2015, **6(1)**, 55-86.
- MEL'ČUK, I., *Language: From Meaning to Text*, Brighton (MA), Academic Studies Press, 2016.

- MEL'ČUK, I., CLAS, A., POLGUÈRE, A., *Introduction à la lexicologie explicative et combinatoire*, Paris/Louvain-la-Neuve, Duculot, 1995.
- MEL'ČUK, I., POLGUÈRE, A., Theory and Practice of Lexicographic Definition, *JOURNAL OF COGNITIVE SCIENCE*, 2018, **19(4)**, 417-470.
- MEL'ČUK, I., POLGUÈRE, A., Les fonctions lexicales dernier cri, in MARENGO, S., *La Théorie Sens-Texte et ses applications. Lexicologie, lexicographie, terminologie, didactique des langues*, coll. « Dixit Grammatica », Paris, L'Harmattan, à paraître.
- MILLER, G. A., BECKWITH, R., FELLBAUM, C. D., Introduction to WordNet : An On-line Lexical Database, *INTERNATIONAL JOURNAL OF LEXICOGRAPHY*, 1990, **3(4)**, 235-244.
- OLLINGER, S., *Le raisonnement analogique en lexicographie, son informatisation et son application au Réseau Lexical du Français*, thèse de doctorat, Nancy, Université de Lorraine, 2014.
- OLLINGER, S., POLGUÈRE, A., Distribution des Systèmes Lexicaux. Version 2.0, Rapport technique, Nancy, ATILF CNRS, 2020a.
- OLLINGER, S., POLGUÈRE, A., CHUDY, Y., GAUME, B., Spiderlex et compagnie, in BENZITOUN, C., BRAUD, C., HUBERT, L., LANGLOIS, D., OUNI, S., POGODALLA, S., SCHNEIDER, S., *6^e conférence conjointe Journées d'Études sur la Parole (JEP, 31^e édition), Traitement Automatique des Langues Naturelles (TALN, 27^e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL, 22^e édition). Volume 4 : Démonstrations et résumés d'articles internationaux*, Nancy, ATALA, 2020, 60–63.
- PAULIN, C., BÉJOINT, H., Introduction. Polysemy: contributions to the grammatical, lexicogenic and lexicographic study of the phenomenon, *LEXIS*, 2008, **1**, 7-9.
- PAUSÉ, M.-S., POLGUÈRE, A. Séparation phraséologique : quand les locutions s'éclatent, *CAHIERS DE LEXICOLOGIE*, 2020, **116**, 233-271.
- PICOCHÉ, J., Polysémie n'est pas ambiguïté, *CAHIERS DE PRAXÉMATIQUE*, 1989, **12**, 75-89.
- PICOCHÉ, J., ROLLAND, J.-C., *Dictionnaire du français usuel. 15.000 mots utiles en 442 articles*, Bruxelles, De Boeck-Duculot, 2002.
- POLGUÈRE, A. Lexical systems: graph models of natural language lexicons, *LANGUAGE RESOURCES AND EVALUATION*, 2009, **43(1)**, 41-55.
- POLGUÈRE, A., Like a Lexicographer Weaving Her Lexical Network, *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex III)*, Mumbai, 15 décembre 2012, 1-3.
- POLGUÈRE, A., From Writing Dictionaries to Weaving Lexical Networks, *International JOURNAL OF LEXICOGRAPHY*, 2014, **27(4)**, 396-418.
- POLGUÈRE, A., Lexical Contextualism: The Abélard Syndrome, in GALA, N., RAPP, R., BEL-ENGUIG, G., *Language Production, Cognition, and the Lexicon*, coll. Text, Speech and Language Technology 48, Cham *et al.*, Springer, 2015, 53-73.

- POLGUÈRE, A., *Lexicologie et sémantique lexicale. Notions fondamentales*, coll. « Paramètres », Montréal, Les Presses de l'Université de Montréal, 2016a.
- POLGUÈRE, A., La question de la géométrie du lexique, *SHS WEB OF CONFERENCES*, 2016b, **27**, 01002.
- POLGUÈRE, A., Il y a un traître par minou : le statut lexical des clichés linguistiques, *CORELA* [En ligne], 2016c, **19**.
- POLGUÈRE, A., A Lexicographic Approach to the Study of Copolysemy Relations, *RUSSIAN JOURNAL OF LINGUISTICS*, 2018, **22(4)**, 788-820.
<http://journals.rudn.ru/linguistics/issue/view/1142>.
- PUSTEJOVSKY, J., *The Generative Lexicon*, Cambridge MA & Londres, The MIT Press, 1995.
- RÉCANATI, F., *Literal Meaning*, Cambridge, Cambridge University Press, 2004.
- SAUSSURE, F. de, *Cours de linguistique générale*, Paris, Payot & Rivage, 1995 [1916].
- SIGMAN, M., CECCHI, G. A., Global organization of the Wordnet lexicon, *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 2002, **99(3)**, 1742-1747.
- SUMMERS, D., *Longman Language Activator. The World's First Production Dictionary*, Harlow, Longman Group, 1993.
- TLFi, *Trésor de la Langue Française informatisé*, Nancy, ATILF CNRS, 2002, URL : www.atilf.fr/tlfi.
- WATTS, D. J., STROGATZ, S. H., Collective Dynamics of 'Small-World' Networks, *NATURE*, 1998, **393**, 440-442.