

Polysémie et homonymie des locutions dans un modèle lexical

Pausé, Marie-Sophie & Sikora, Dorota

Université de Lorraine & ATILF-CNRS, Université du Littoral Côte d'Opale & HLLI
marie-sophie.pause@univ-lorraine.fr
dorota.sikora@univ-littoral.fr

Résumé. Si les unités phraséologiques, telles que les locutions qui nous intéressent dans cet article, sont aujourd'hui de plus en plus régulièrement examinées dans les ressources lexicographiques (cf. WordNet[1], FrameNet[2] à titre d'exemple), leur traitement et la description qui en est proposée prennent rarement en compte de manière systématique leur polysémie. La base de données lexicales *Réseau Lexical du Français (RL-fr)*, développée au laboratoire ATILF – CNRS (UMR 7118)[3], compte actuellement environ 3000 entrées locutionnelles. En nous appuyant sur les recherches menées dans le cadre théorique de la *Lexicologie Explicative et Combinatoire* (Mel'čuk et al., 1995) et sur la pratique du travail de description lexicographique liée au développement de la base *RL-fr*, nous nous intéressons aux spécificités de la polysémie des vocables locutionnels et à leur traitement dans une ressource dictionnaire à large couverture. Après une présentation générale du *RL-fr*, nous discutons l'analyse et l'encodage des locutions à sens multiples. En effet, pour un vocable locutionnel tel que *marcher sur la tête* qui nous sert d'exemple, trois sens sont à considérer : 'dominer', 'remporter une victoire écrasante' et 'se comporter de manière déraisonnable'. L'analyse consiste alors à déterminer si nous sommes face à un seul vocable polysémique ou bien s'il s'agit d'homonymes, et par conséquent de deux, voire trois entrées de dictionnaire distinctes. Le modèle d'analyse développé et pratiqué dans le cadre du *RL-fr* prend en compte aussi bien les liens sémantiques entre les (éventuelles) acceptions copolysèmes d'une locution, que les aspects formels de son signifiant, à travers une description de sa structure lexico-syntaxique. Cette dernière s'avère plus d'une fois décisive, lorsqu'il faut juger s'il est question de polysémie ou d'homonymie. Ainsi, l'analyse de la structure lexico-syntaxique de *marcher sur la tête* nous amène à constater qu'il existe un vocable polysémique *marcher sur la tête*¹ regroupant les acceptions 'dominer' et 'infliger une grosse défaite' d'une part, et de l'autre son homonyme *marcher sur la tête*² avec un seul sens, celui de 'se comporter de manière déraisonnable'. En conclusion, nous considérons qu'une description de la structure lexico-syntaxique est fondamentale, si l'on tient à distinguer des cas de polysémie de ceux d'homonymie des locutions. Nous allons même jusqu'à dire que la structure lexico-syntaxique d'une locution constitue le signifiant de ces unités phraséologiques.

Abstract. The present paper focuses on a model of linguistic analysis that enables lexicographers to draw a clear distinction between polysemy and homonymy of idioms. Indeed, researchers' growing interest in phraseological units leads to give more room to idioms in lexicographic description (see for instance WordNet[1], FrameNet[2]). We therefore believe that, few dictionaries—if any—systematically account for their polysemy. The lexical database *Réseau Lexical du Français (French Lexical Network, henceforth RL-fr)* currently developed in the laboratory ATILF – CNRS (UMR 7118)[3] presently encodes about 3000 French idioms, with a systematic treatment of their polysemous structure. Building on this large-scale lexicographic

work based on theoretical et methodological principles of *Explanatory and Combinatorial Lexicography* (Mel'čuk et al. 1995), we examine the case of the verbal idiom *marcher sur la tête* (lit. 'to walk on the head'). It conveys three senses to be described in a dictionary: 'to dominate', 'to inflict a crushing defeat (in a competition)', and 'to behave irrationally'. Thus, the question that arises is whether there is one polysemous idiom *marcher sur la tête* or if we should rather consider it a case of homonymy. Semantic analysis can hardly establish any meaning relation between 'to dominate' and 'to inflict a crushing defeat' on the one side, and 'to behave irrationally' on the other. However, lexicographic treatment of idioms in *RL-fr* takes into account the signifié as well as the signifiant: the form of the later is modeled as its lexico-syntactic structure. As a matter of fact, *marcher sur la tête* reveals two different lexico-syntactic structures, corresponding respectively to the polysemous idiom *marcher sur la tête*¹ with two meanings ('to dominate', 'to inflict a crushing defeat') and its homonym *marcher sur la tête*²—'to behave irrationally'. These results lead to conclude that formal properties of idioms, modeled as their lexico-syntactic structure, should systematically be accounted for in lexicographic description. Moreover, given the available data, we suggest that lexico-syntactic structures of idioms correspond to their signifiants.

[1] <https://wordnet.princeton.edu/>

[2] <https://framenet.icsi.berkeley.edu/fndrupal/>

[3] www.atilf.fr

1 Introduction : questions de recherche et objectifs de l'article¹

La polysémie des locutions, syntagmes sémantiquement non compositionnels, est-elle différente de celle des vocables lexémiques ? Si oui, où résident les différences ? Dans les dictionnaires traditionnels, les locutions ne sont généralement pas considérées comme des unités lexicales à part entière et reçoivent une description hétérogène (Pruvost, 2011)². Seules des ressources récentes, adossées à des recherches linguistiques, telles que *WordNet* (Fellbaum, 1998, Osherson et al., 2009), *FrameNet* (Fillmore et al., 2003), et le *Réseau Lexical du Français* (voir *infra*) leur accordent ce statut. Malgré un intérêt grandissant des chercheurs pour la phraséologie, peu d'études abordent la polysémie des locutions, notamment en français. Notons qu'il existe un numéro de *Syntaxe et Sémantique* intitulé « Polysémie et Polylexicalité » (2004), mais les unités discutées sont majoritairement des mots composés, ou bien des unités telles que TELEPHONE PORTABLE, dont le statut lexical varie entre locution nominale et nom composé³.

S'appuyant sur les travaux et la pratique lexicographique liés au développement de la base de données *Réseau Lexical du Français (RL-fr)*, cf. Gader et al., 2012, Polguère, 2014), cet article s'intéresse aux spécificités de la polysémie des vocables locutionnels et à son traitement lexicographique dans le *RL-fr*. La polysémie – on le sait – c'est le fait, pour un même signifiant, de recouvrir plusieurs signifiés sémantiquement liés. Le *RL-fr* permet de traiter et de décrire ces deux facettes – la forme et le sens – tout en prenant en compte les propriétés combinatoires des locutions.

Nous commencerons, dans la section 2, par une rapide présentation de la ressource et de la notion de *système lexical*, c'est-à-dire du modèle du lexique dont la ressource constitue l'implantation. La partie suivante sera consacrée à la discussion de l'analyse et de l'encodage de la polysémie des locutions. Nous nous arrêterons ensuite, en section 4, sur la façon dont le *RL-fr* traite leurs caractéristiques formelles, en associant à chacune d'entre elles une structure lexico-syntaxique, qui enrichit – d'une manière inédite à notre connaissance – la description lexicographique des unités polylexicales. On verra que la structure lexico-syntaxique d'une locution peut se montrer décisive pour distinguer entre la polysémie et l'homonymie. Nous étudierons le cas de MARCHER SUR LA TÊTE et des trois acceptions qu'on lui connaît : leur traitement requiert, d'une part, la prise en compte de relations sémantiques entre les signifiés, et de l'autre, l'analyse de la structure des signifiants. Nous concluons nos réflexions dans la cinquième et dernière section.

2 Présentation de la base lexicale Réseau Lexical du Français

Du point de vue de ses utilisateurs, la base de données *RL-fr* développée au sein du Laboratoire ATILF – CNRS (UMR 7118)⁴ est une ressource de type dictionnaire virtuel (Selva et al., 2003). Les informations qu'elle réunit peuvent être explorées avec des visées applicatives allant de l'enseignement de langues à la génération automatique de dictionnaires de spécialité. Dans une perspective plus large, le *RL-fr* est une implantation d'un modèle du lexique français⁵ élaboré selon les principes théoriques et méthodologiques homogènes, définis initialement dans le cadre de la Lexicologie Explicative et Combinatoire. Ce modèle a été récemment développé par Polguère (2009, 2014) sous le nom de *système lexical*.

Les propriétés formelles d'un système lexical ont été exposées et discutées entre autres dans Polguère (2014), puis dans Ollinger (2014). Faute de place, nous nous limiterons ici à relever celles d'entre elles qui sont particulièrement importantes pour l'intégration et la description de la polysémie et de l'homonymie dans le *RL-fr*.

Comme son nom l'indique, le *RL-fr* est structuré en réseau et forme un graphe au sens mathématique du terme : il est fait de sommets (nœuds) que constituent essentiellement des lexies (unités lexicales, qui sont soit des lexèmes, soit des locutions) et des arcs représentant des relations lexicales de différente nature (cf. *infra*). Précisons qu'il s'agit d'un modèle non-ontologique, puisque tous les liens ont un statut égal. C'est dire que contrairement à une ressource telle que *WordNet*, le *RL-fr* ne privilégie d'aucune façon les relations qui renvoient à une organisation conceptuelle du lexique (cf. Polguère, 2014).

La figure 1 présente le graphe *RL-fr* vu « de l'extérieur », une photo d'identité en quelque sorte, qui permet de remarquer ses traits caractéristiques, c'est-à-dire ses propriétés structurelles essentielles.

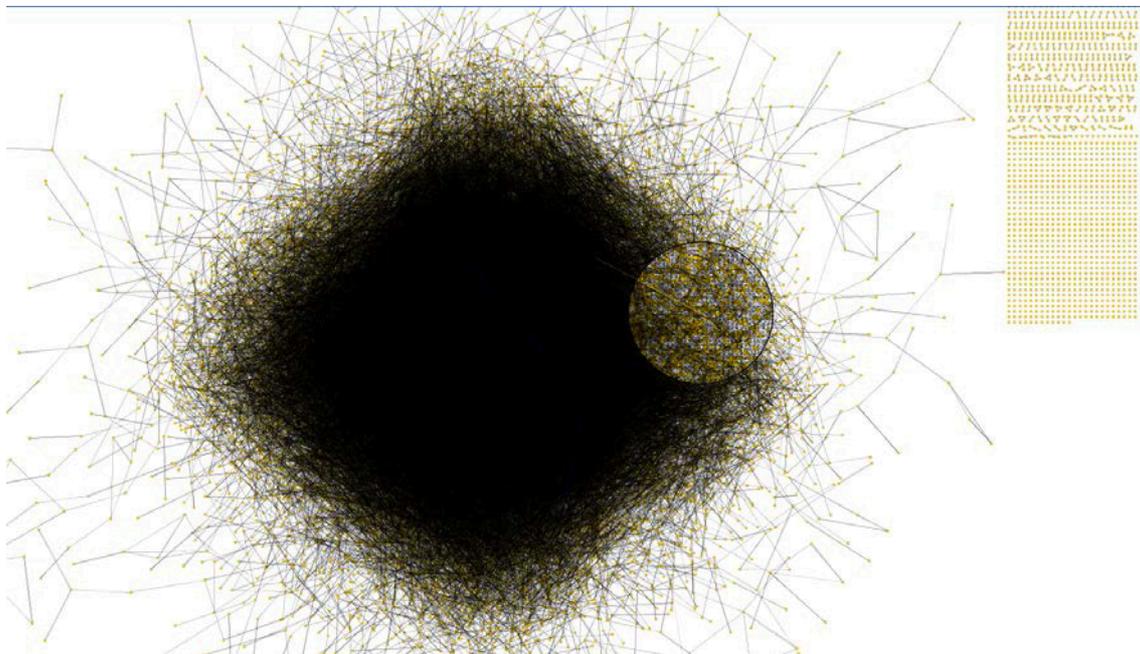


Figure 1 : Graphe *RL-fr* (état de développement du 23 novembre 2015).

Dans la mesure où il s'agit d'une ressource en cours de développement, la structure du graphe est en constante évolution. Au fur et à mesure que le travail des lexicographes progresse, de nouveaux sommets

y sont insérés et un réseau de liens lexicaux de plus en plus dense se constitue, en rattachant les lexies faiblement, voire pas du tout connectées, à l'ensemble du graphe. À terme, chaque sommet sera rattaché à au moins un autre nœud, dans une structure de type graphe petits mondes. Les propriétés formelles des graphes petits mondes ont fait objet de nombreuses publications récentes (Gaume 2004, et surtout Ollinger 2014 pour une discussion relative au *RL-fr*). Notons simplement avec Kiss (1968) que leur organisation est à l'image de celle du lexique d'une langue. Ainsi, malgré une connectivité généralement faible de un ou deux arcs pour la plupart des sommets, il est possible de naviguer entre eux, en passant d'un nœud vers n'importe quel autre en un nombre de pas peu élevé.

Les sommets encapsulent une description lexicographique comprenant les caractéristiques grammaticales de la lexie-vedette, des informations relatives à son sens (spécification de ses actants pour les unités à caractère prédicatif, étiquette sémantique⁶ et, pour certaines, la définition⁷), les relations lexicales dont elle est source et cible, ainsi que des citations illustrant ses emplois. Au lieu de former un bloc, l'entrée lexicale est divisée en zones allouées à chaque type d'informations. Cette organisation non-atomique des données, commune à tous les sommets qu'ils soient lexémiques ou locutionnels (cf. figures 2 et 3 ci-dessous) permet une exploration « thématique » de la base.

[CC]	nom commun fém
[MO]	Modèle : Noms, Table : chat sing : voiture plur : voitures
[DF]	véhicule voiture qui sert à X_{-1} pour Y_{-2}
[PL]	Syn ₁ : berline ₁ , break ² ₁ , calèche, carrosse, voit coche, diligence, coupé ₁ , sulky, tilbury ; fiacre ; carriole, chariot, charrette, tombereau Syn ₂ : attelage ₁ Professionnel X S ₁ ^{actant} : cocher fonctionnel Mero : brancard, essieu, roues, marche pied Instr : en _{traj} ₁ [-] Loc ₁ : à bord ₁ [99 ART -], dans ₁ [ART -] [X] faire fonctionner - Real ₁ : conduire ₁ [ART -] [Y] se mettre en position d'utiliser - Real ₂ ¹ : monter ₁ [3 en -] [Loc_in ART -], prendre place [Loc_in ART -] Real ₂ ² : rouler ¹ ₁ [3 en -, Loc_in ART -]; se déplacer ₁ [3 en -, voyager [en -, Loc_in ART -]; se promener [en -] Real ₂ ³ : descendre ₁ [3 en ART -] Real ₂ ⁴ : tirer ₁ [2 ART -] PreparReal ₁ : atteler [ART N=33 à ART -] LiquReal ₁ : dételé [ART N=33 09 ART -] déplacement Fact ₁ : rouler ¹ ₁ [2a Fact ₁ : transporter ₁ [N-V]
[EX]	Nous possédions alors plusieurs voitures tirées par des chevaux, et nous donnions des représentations théâtrales et des pantomimes avec deux hommes qui étaient les cousins de ton père et mes deux sœurs. Frantext DHOTEL André, <i>Le Pays où l'on n'arrive jamais</i> , 1955, p. 242

Figure 2 : Informations stockées dans un nœud lexémique (lexème VOITURE II).



Figure 3 : Informations stockées dans un nœud locutionnel (CHEVEUX D'ANGE II).

Les sommets sont reliés par des arcs qui, grâce au système de fonctions lexicales (Mel'čuk, 1996), modélisent les relations lexicales suivantes :

1. Relations paradigmatiques : comme toute autre lexie, une locution possède des synonymes, des antonymes et des hyper- ou des hyponymes. Ainsi, la locution nominale CHEVEUX D'ANGE II est reliée à son quasi-synonyme VERMICELLE (et *vice versa*) et à son hyperonyme PATE I.1. De plus, les relations actanciennes sont systématiquement encodées de manière à connecter une lexie telle que CARTON ROUGE I.1, 'objet servant à transmettre une information', aux noms de ses trois actants, respectivement : ARBITRE_{N,masc} et ARBITRE_{N,fém}, JOUEUR 2 et JOUEUSE 2, et FAUTE I.1b.
2. Relations syntagmatiques, qui pointent vers les collocatifs formant la combinatoire restreinte de la lexie vedette. Dans le cas de CARTON ROUGE I.1, il s'agira d'indiquer entre autres les verbes qui dénotent l'usage qui peut en être fait : SORTIR II.1a, BRANDIR, DEGAINER.

Un système lexical tel que le *RL-fr* permet de tisser tout un ensemble d'autres liens aussi bien entre des sens qu'entre des formes. Ainsi, les lexies-acceptations d'un vocable polysémique sont connectées entre elles par des liens de copolysémie. Le vocable en question forme un graphe, représentant l'hypothèse que fait le lexicographe sur son sens de base en synchronie et sur les ponts sémantiques (Mel'čuk et al., 1995) – condition *sine qua non* de toute polysémie – entre les lexies copolysèmes. La figure 4 illustre le cas de VOITURE, avec son réseau d'acceptations copolysèmes tel qu'il s'affiche dans le *RL-fr*. Nous y reviendrons plus amplement dans la section suivante, en en proposant une version enrichie d'exemples.

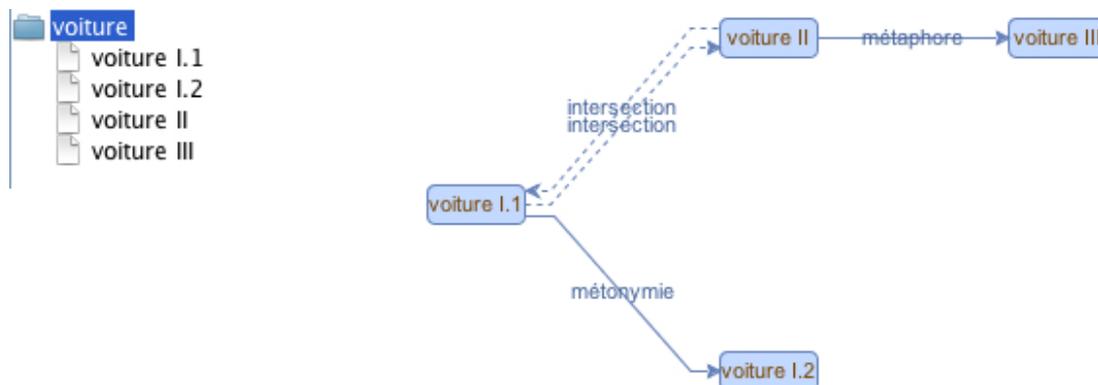


Figure 4 : Réseau des copolysèmes du vocable VOITURE.

Le travail d'élaboration des paraphrases définitoires lancé depuis peu conduit les lexicographes à tisser des liens d'inclusion définitionnelle ; à partir du *definiens* de la lexie CIGARETTE I, on atteindra les lexies essentielles pour la compréhension de ce sens : FUMER III.a, TABAC I.2, LEVRE I.1 et DOIGT I.1a (cf. figure 5).

[DF]
chose fabriquée pour être fumée
cigarette destinée à être fumée par X_{-1}
=
objet I.1a destiné à être fumé III.a par l'individu X
de forme allongée
constitué d'un petit tube de papier très fin rempli de tabac¹ I.2 haché, que X fait se consumer par une de ses extrémités que X tient entre ses lèvres I.1 ou entre ses doigts I.1a quand X fume

Figure 5 : Encodage de la définition de CIGARETTE I.

De plus, quel que soit leur statut (locutions, clichés linguistiques, proverbes, etc.), les unités polylexémiques se rattachent au graphe *RL-fr* par des liens d'inclusion formelle qui leur sont spécifiques. À ce stade, nous ne faisons que les mentionner pour compléter la description générale de la ressource, en remettant leur présentation à la section 4.

3 Polysémie des locutions dans le Réseau Lexical du Français

Dans la présente section, nous exposons la façon dont le *RL-fr* décrit la pluralité des sens. En 3.1, nous présentons l'approche générale de la polysémie de la ressource. Nous observerons ensuite, en 3.2, les liens de copolysémie des locutions.

3.1 Traitement de la polysémie dans le Réseau Lexical du Français

Pour organiser la polysémie d'un vocable, le *RL-fr* adopte une approche dérivationnelle synchronique. Dans la figure 6 ci-dessous, nous reproduisons le réseau des copolysèmes du vocable VOITURE, en y intégrant des exemples illustrant chaque acception.

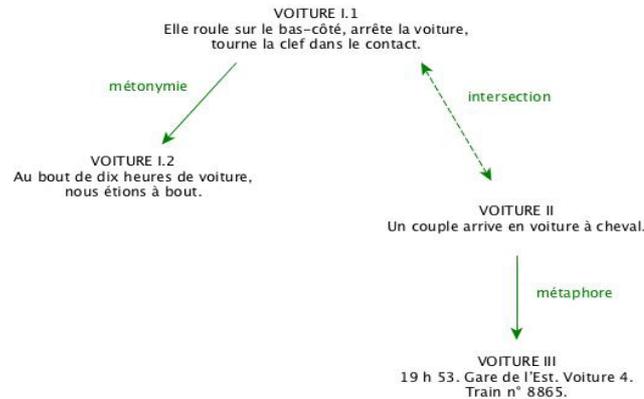


Figure 6 : Réseau des copolysèmes du vocable VOITURE (avec illustration de sens).

Si nous qualifions cette approche de dérivationnelle, c'est parce qu'elle consiste à organiser les acceptions à partir d'un sens de base (unité lexicale de base, désormais ULB). C'est l'ULB qui permet de dériver les autres copolysèmes, c'est-à-dire d'identifier et de tisser des liens sémantiques directs ou indirects entre les copolysèmes. La notion de *pont sémantique* proposée dans Mel'čuk, Clas et Polguère (1995 : 157) se trouve ainsi formellement encodée.

L'ULB d'un vocable est identifiée, nous le rappelons, dans une perspective synchronique : dans le cas d'un vocable tel que VOITURE, le graphe est organisé à partir du sens 'véhicule automobile', et non pas VOITURE II 'véhicule monté sur roues, tiré ou poussé par un animal, un homme'⁸, même si ce sens est antérieur d'un point de vue diachronique. On considère que, pour les locuteurs du français contemporain, l'ULB est à la fois le sens cognitivement le plus saillant parmi les copolysèmes (cf. Wilkins, 1996) et sémantiquement le plus simple à formuler. Cette saillance peut être facilement testée par une simple question adressée à un locuteur : *Qu'est-ce ce que veut dire VOITURE ?* Il y a fort à parier qu'en dehors d'un contexte spécifique, une réponse spontanée consistera à paraphraser le sens de VOITURE I.1, 'véhicule automobile' plutôt que celui de VOITURE II.

La métonymie qui relie l'acception VOITURE I.2, 'fait de se déplacer en voiture', à l'ULB et la relation métaphorique entre VOITURE II et VOITURE III sont des mécanismes bien connus dans le développement de la polysémie. Ils ne permettent cependant pas de modéliser la relation – clairement perceptible même pour un locuteur peu au fait des évolutions diachroniques – entre l'ULB et VOITURE II. Dans les deux cas, il s'agit de 'véhicule' ou 'moyen de transport' : cette relation sémantique particulière qui repose sur une partie commune aux deux sens est saisie en tant qu'intersection.

En terminant cette présentation très, et faute de place trop, rapide du traitement de la polysémie dans le *RL-fr*, notons que les ponts sémantiques (les arcs représentés en vert dans la figure 6) devront à terme être explicités dans les définitions des lexies copolysèmes qui en sont la source et la cible. Autrement dit, les définitions devront être formulées de manière à faire apparaître les relations de copolysémie. Il se peut donc qu'un graphe de copolysèmes dégagé à l'issue d'une première analyse soit remodelé lors de l'élaboration des paraphrases définitoires.

3.2 Liens de copolysémie des locutions

Dans l'état actuel du développement du *RL-fr*, la modélisation des relations de copolysémie autorise des comparaisons entre les structures de polysémie des vocables lexémiques et locutionnels. D'une part, il est possible d'observer certaines régularités, et de l'autre, de repérer nombre de difficultés liées

spécifiquement à la polysémie des locutions. Deux points, mis en évidence grâce à l'encodage systématique des liens de copolysémie, méritent d'être soulignés.

Premièrement, on note d'importantes différences entre le taux de polysémie des vocables lexémiques et celui des vocables locutionnels. Pour l'ensemble de la nomenclature, soit 26 540 vocables et 17 262 unités lexicales, ce taux se situe aux alentours de 1,54. Il est de 1,62 pour les vocables lexémiques et de 1,14 seulement en ce qui concerne ceux qui sont des locutions. De manière générale, même dans le cas de locutions relativement fréquentes, les réseaux de copolysèmes ont moins de sommets que ceux de vocables lexémiques fréquents. Leur taux de polysémie est généralement plus faible, dans la mesure où le nombre d'acceptions est rarement supérieur à trois. Les réseaux de copolysèmes lexémiques peuvent au contraire former des graphes atteignant, pour les plus fréquents d'entre eux, une vingtaine de sommets (cf. par exemple le verbe PASSER regroupant 19 copolysèmes⁹, et TETE avec 24 acceptions décrites à ce jour).

Deuxièmement, une comparaison des graphes que forment les acceptions copolysèmes des vocables lexémiques et locutionnels met en avant des similarités dans les structures dérivationnelles. Les arcs indiquant la nature des relations entre les copolysèmes montrent, en effet, que les mêmes mécanismes (métaphore, métonymie ou extension) sont en œuvre. Les locutions semblent donc soumises aux mêmes modes de fonctionnement cognitif que ceux qui caractérisent les lexèmes dans la plupart, sinon dans toutes les langues (cf. Sikora, 2014).

La figure 7 ci-dessous permet de comparer la polysémie de MARATHON et de COURSE D'OBSTACLES. Leurs graphes sont relativement simples, puisque ces vocables ne comptent que deux acceptions, dont la première est la source d'un développement métaphorique.

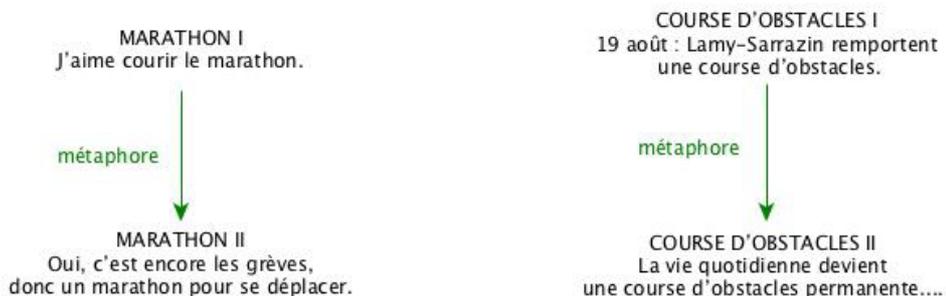


Figure 7 : Graphes de copolysèmes d'un vocable lexémique (MARATHON) et d'un vocable locutionnel (COURSE D'OBSTACLES).

Les liens de copolysémie permettent de traiter le versant sémantique de la polysémie, c'est-à-dire d'explicitier les relations de sens entre les signifiés d'un vocable lexémique ou locutionnel. Cependant, tout au long de notre travail, nous avons repéré des vocables locutionnels pour lesquels ce type de connexions est problématique à établir, tout comme il peut, par ailleurs, l'être pour certains lexèmes que l'on doit considérer dès lors comme homonymes. Il en est ainsi dans le cas de MARCHER SUR LA TETE illustré dans les exemples (1) à (3) ci-dessous. Si le lien métaphorique entre l'acception de la phrase (1), 'avoir un comportement dominant à l'égard de quelqu'un', et celle en (2), 'infliger une lourde défaite à quelqu'un', est facile à identifier, il n'en va pas de même pour MARCHER SUR LA TETE de l'exemple (3), que l'on pourrait paraphraser par 'avoir un comportement déraisonnable'.

1. La fille, surprise d'un tel débordement verbal, eut son premier sourire et murmura en s'agaçant l'oreille d'une tige de graminée : – Vous pourriez vivre avec une jeune fille qui vous **marcherait sur la tête** ? (Frantext)

2. L'équipe US version 2002 a été humiliée trois fois pour ne finir qu'à la sixième place, le plus mauvais rang des États-Unis en quatorze éditions. Dans l'ordre, l'Argentine en poule (87-80), la Yougoslavie en quart de finale (81-78) et l'Espagne en match de classement (81-75) lui **ont marché sur la tête**. (*L'Est Républicain* 10 septembre 2002)
3. Les hommes politiques **marchent sur la tête** ! Il faudrait qu'on m'explique comment on peut recapitaliser la BEI avec de l'argent qu'on n'a pas ? (Web)

Du côté du signifié, le pont sémantique entre la locution MARCHER SUR LA TÊTE dans (3) et celles des exemples (1) et (2) semble pour le moins problématique. Nous allons voir à présent que les liens d'inclusion formelle mis en évidence par la structure lexico-syntaxique des phrasèmes apportent des informations supplémentaires sur le signifiant, en permettant de trancher efficacement entre polysémie et homonymie.

4 Structures lexico-syntaxiques des locutions

Dans cette section, nous nous tournons vers le signifiant des locutions. En tant qu'unités polylexicales syntagmatiques, les locutions ont des constituants lexicaux, liés par des relations syntaxiques. Cette structure de leurs signifiants est également décrite dans le *RL-fr*. Chaque locution est alors liée à ses constituants lexicaux, et au patron syntaxique sur lequel elle est formée. Le résultat obtenu est une structure lexico-syntaxique (SLS). Dans la section 4.1, nous présentons rapidement l'intérêt qu'elle présente pour une analyse linguistique.

Précisons toutefois que l'identification des SLS n'est pas liée directement au traitement de la polysémie. Il n'en est pas moins que, chemin faisant, ce travail permet de porter un regard nouveau sur la polysémie des locutions. La section 4.2 discute des cas où le recours aux SLS conduit à une description plus précise de la pluralité des sens, notamment lorsqu'il fournit des moyens de décider s'il s'agit de polysémie (cf. MARCHER SUR LA TÊTE des exemples 1 et 2 de la section 3.2 ci-dessus) ou s'il faut opter pour l'homonymie.

4.1 Structures lexico-syntaxiques et liens d'inclusion formelle

L'objectif premier du travail de description lexico-syntaxique des locutions est de proposer modélisation, basé sur la syntaxe de dépendance, permettant d'expliquer les variations formelles des locutions (Pausé, à *paraître*). À terme, chaque locution sera, au sein de son article lexicographique, liée à un arbre de dépendance contenant des informations indiquant pourquoi certaines variations syntagmatiques se trouvent bloquées, alors que d'autres – passivation (4a), détachement (4b), extraction (4c), relativation (4d) et ajout d'un conjoint sur une composante autre que la tête du syntagme (4e) – sont autorisées (Pausé, *en préparation*).

4.
 - a. Bon, maintenant que **la glace est brisée** et que nous avons fait connaissance, [...] vous allez peut-être, monsieur Cidrolin qui semblez être du quartier, nous indiquer un autre endroit pas trop éloigné du centre de la ville, où mes chevaux [sic], ma fille, Empoigne, sa maman et moi pourrions combiner les joies du camping [sic] avec les plaisirs de la capitale. (Frantext)
 - b. Finalement, Régimbald a eu sa chance. **La glace, il l'a brisée** le 17 août dernier, et pour ajouter au stress d'un premier match dans la Ligue canadienne de football, c'est contre les Alouettes de Montréal qu'il a fait ses premiers pas. (Web)
 - c. Si on veut retrouver la confiance et le respect du peuple, **c'est devant sa porte qu'il faut** commencer à **balayer**. (Web)
 - d. Tant mieux pour Cavada, qu'il se prenne dans les gencives **la veste qu'il n'arrête pas de retourner** ! Ce type-là est à vomir ! (FrWac)
 - e. Monsieur le Président, Monsieur le Président du Conseil, vous devez **prendre par les cornes le taureau du doute sur Lisbonne** ; la France a déjà une longue

expérience dans l'art de tantôt faire avancer, tantôt faire reculer la construction européenne. (Web)

Une étude des variations paradigmatiques est également envisagée, dans le but de décrire systématiquement les commutations possibles (5) ou impossibles (6).

5.
 - a. [...] [L]e SLUC Nancy peut-il croire en ses chances de succès ce soir face à une équipe de Pau-Orthez taillée dans le roc et qui lui a fait **manger la poussière** à quatre reprises cette saison, championnat et coupe confondus ? (*L'Est Républicain* 28 mai 2002)
 - b. On veut une brute, un vrai leader qui fera **bouffer la poussière** à nos ennemis [...]. (Web)
6.
 - a. Comme beaucoup d'autres, mes parents ont adhéré très jeunes, dix-huit ans à peine, à un discours libérateur : ils en **ont payé les pots cassés**. Ils n'avaient pas fait d'études, ce n'étaient pas des intellectuels, ils avaient rompu avec leur famille... (Frantext)
 - b. *Ils en ont réglé les pots cassés.

À ce jour, plus de 2500 locutions françaises ont reçu une SLS dans le *RL-fr*. Les figures 8 et 9 montrent qu'une locution peut contenir soit uniquement des lexèmes, soit des lexèmes et des locutions.

The screenshot shows a software interface for analyzing the locution 'manger les pissenlits par la racine'. The main window displays the locution and its classification as 'locution verbale' and 'locution forte'. A detailed view of the 'Structure syntaxique' is shown, listing components with their grammatical categories, forms, and probabilities.

Composant	Acception	Forme	Probabilité
V	manger _v I.1a		100
Art	le _{Art}	les	100
NC	pissenlit	pissenlits	100
Prép	par III	par	100
Art	le _{Art}	la	100
NC	racine I		100

Figure 8 : SLS de la locution verbale MANGER LES PISSENLITS PAR LA RACINE 'être mort'.

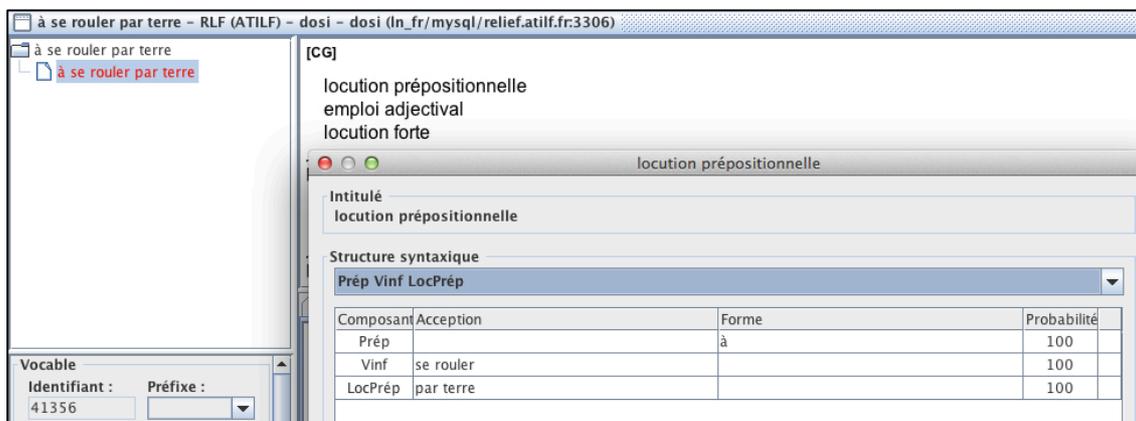


Figure 9 : SLS de la locution prépositionnelle A SE ROULER PAR TERRE¹⁰ ‘très drôle’.

La démarche d’identification des SLS permet de réunir des éléments expliquant leur comportement, tout en développant progressivement une base de patrons syntaxiques qui à terme devrait couvrir l’ensemble des locutions françaises.

L’intérêt d’identifier les constituants lexicaux des locutions est lié au fait que la combinatoire des constituants a un impact sur la combinatoire des locutions elles-mêmes. Nous savons par exemple que dans les syntagmes libres (exemples en 7), certaines lexies permettent la montée du possesseur, qui se manifeste à travers la pronominalisation du possesseur. C’est le cas des lexies désignant des parties du corps (7a) ou d’objets (7b), ou encore des vêtements (7c).

7.
 - a. Jeanne lui a cassé le bras, à Line.
 - b. Je lui ai abîmé une patte, à la chaise.
 - c. Je lui ai abîmé une manche, à ce gilet.

Cette montée du possesseur par un pronom personnel complément est également observable dans le cas de certaines locutions, par exemple pour MARCHER SUR LA TÊTE dans (1) et (2). À l’inverse, elle ne peut s’admettre pour MARCHER SUR LES TRACES ‘suivre l’exemple’ (8), puisque TRACE n’est pas liée à son actant par un lien de possession inaliénable. Elle est également impossible dans le cas de (3), étant donné que la locution a un unique actant qui est son sujet (ce cas est discuté plus précisément en section 4.2 *infra*).

- (1) La fille, surprise d’un tel débordement verbal, eut son premier sourire et murmura en s’agaçant l’oreille d’une tige de graminée : – Vous pourriez vivre avec une jeune fille qui **vous marcherait sur la tête** ?
 - (2) L’équipe US version 2002 a été humiliée trois fois pour ne finir qu’à la sixième place, le plus mauvais rang des États-Unis en quatorze éditions. Dans l’ordre, l’Argentine en poule (87-80), la Yougoslavie en quart de finale (81-78) et l’Espagne en match de classement (81-75) **lui ont marché sur la tête**.
 - (3) Les hommes politiques **marchent sur la tête** ! Il faudrait qu’on m’explique comment on peut recapitaliser la BEI avec de l’argent qu’on n’a pas ? (Web)
8.
 - a. À l’occasion des championnats académiques 2002, organisés cet après-midi à Besançon, les participants essaieront ainsi de **marcher sur les traces de leurs glorieux aînés**. (L’Est Républicain 11 décembre 2002)

- b. *Les participants essaieront de **leur marcher sur les traces**.

L'intérêt d'identifier le patron syntaxique des locutions est lié à l'hypothèse selon laquelle la structure syntaxique d'une locution a un impact sur son emploi. Ainsi, une locution construite à partir d'un verbe transitif tel que FRAPPER dans (9) ou di-transitif (PRENDRE dans l'exemple 10) sera plus encline à la passivation.

9. **Un grand coup a été frappé**, hier, à la salle polyvalente, lors de la première journée des Rencontres internationales de percussions. (Web)
10. Nous allions voir, ce que nous allions voir, **le taureau serait pris par les cornes**, demain ne serait plus comme hier. (FrWac)

Il nous est impossible de discuter, dans le cadre de cet article portant sur la polysémie des locutions, de toutes les applications possibles des SLS. Avant de quitter cette section, nous tenons néanmoins à souligner que l'identification des lexies formellement incluses dans les locutions n'est en rien incompatible avec leur non-compositionnalité sémantique. Une locution telle que MARCHER SUR LA TÊTE de l'exemple (1) est non compositionnelle dans la mesure où, en réponse à la question de savoir quel est son sens, un locuteur contemporain du français répondra plus que probablement « *Marcher sur la tête de quelqu'un*, cela signifie avoir un comportement dominant à son égard ». Une définition lexicographique du *RL-fr* reprendra cette explication dans une paraphrase construite dans un langage formel normé. L'identification des lexies incluses ne fait pas partie de l'analyse décompositionnelle du sens. Elle consiste à explorer une plus ou moins forte transparence du sens de la locution pour retrouver les éléments qui l'ont motivée. Cela revient donc à répondre à la question « Pourquoi dit-on de quelqu'un qui a un comportement dominant à l'égard d'un autre individu qu'*il lui marche sur la tête* ? ». Dans ce travail, le lexicographe étudie donc l'origine des locutions, telle qu'elle est attestée par les étymologistes, ou bien reconstituée par la communauté linguistique. Il se situe ainsi dans des perspectives que Perrin (2014) qualifie respectivement de versants historique et ascendant des sens locutionnels, par opposition au versant descendant, qui est le sens global – non compositionnel – de la locution.

4.2 Polysémie et/ou homonymie : ce que peuvent révéler les structures lexico-syntaxiques

En commentant les exemples (1) à (3) dans la section 3.2, nous avons signalé la difficulté d'identifier la relation sémantique, *ipse eo* de construire un pont sémantique, entre les sens de la locution MARCHER SUR LA TÊTE dans les phrases (1) et (2) d'une part, et (3) de l'autre. Cela revient donc à constater qu'une analyse décompositionnelle des signifiés conduit vers l'hypothèse de deux vocables locutionnels homonymes. Rappelons ces exemples avant de poursuivre notre discussion.

- (1) La fille, surprise d'un tel débordement verbal, eut son premier sourire et murmura en s'agaçant l'oreille d'une tige de graminée : – Vous pourriez vivre avec une jeune fille qui vous **marcherait sur la tête** ? (Frantext)
- (2) L'équipe US version 2002 a été humiliée trois fois pour ne finir qu'à la sixième place, le plus mauvais rang des États-Unis en quatorze éditions. Dans l'ordre, l'Argentine en poule (87-80), la Yougoslavie en quart de finale (81-78) et l'Espagne en match de classement (81-75) lui **ont marché sur la tête**. (*L'Est Républicain* 10 septembre 2002)
- (3) Les hommes politiques **marchent sur la tête** ! Il faudrait qu'on m'explique comment on peut recapitaliser la BEI avec de l'argent qu'on n'a pas ? (Web)

Observons à présent les SLS correspondant aux syntagmes que forment ces locutions. On constate en effet deux patrons syntaxiques différents que l'on associe à MARCHER SUR LA TÊTE des l'exemples (1) et (2) d'une part, et (3) d'autre part. Comparons les figures 10 et 11 :

Composant	Acception	Forme
V	marcher III	
Prép		sur
Art		la
NC	tête I.1a	
Prép		
\$2		

Figure 10 : SLS de la locution MARCHER SUR LA TÊTE 'avoir un comportement dominant à l'égard de quelqu'un', et 'infliger une lourde défaite à quelqu'un' (exemples 1 et 2).

Composant	Acception	Forme	Probabilité
V	marcher I.1a		100
Prép		sur	100
Art		la	100
NC	tête I.1a		100

Figure 11 : SLS de la locution MARCHER SUR LA TÊTE 'avoir un comportement déraisonnable à propos de quelque chose' (exemple 3).

Les SLS montrent que dans les deux cas, les locutions incluent des constituants lexicaux différents. MARCHER SUR LA TÊTE, 'avoir un comportement dominant à l'égard de quelqu'un' est relié à la lexie MARCHER III dont le sens peut être paraphrasé par 'poser son pied sur quelque chose' (cf. figure 10). La structure actancielle de cette lexie comprend deux actants : X marche sur Y. Il en est autrement pour la locution MARCHER SUR LA TÊTE qui signifie 'avoir un comportement déraisonnable' (cf. figure 11), qui renvoie à la lexie MARCHER I.1a 'se déplacer en faisant des pas', qui n'a qu'un seul actant X. Cette différence de constituants se traduit par deux systèmes de dépendances différents pour les locutions qui les contiennent. La figure 12 permet de les comparer.

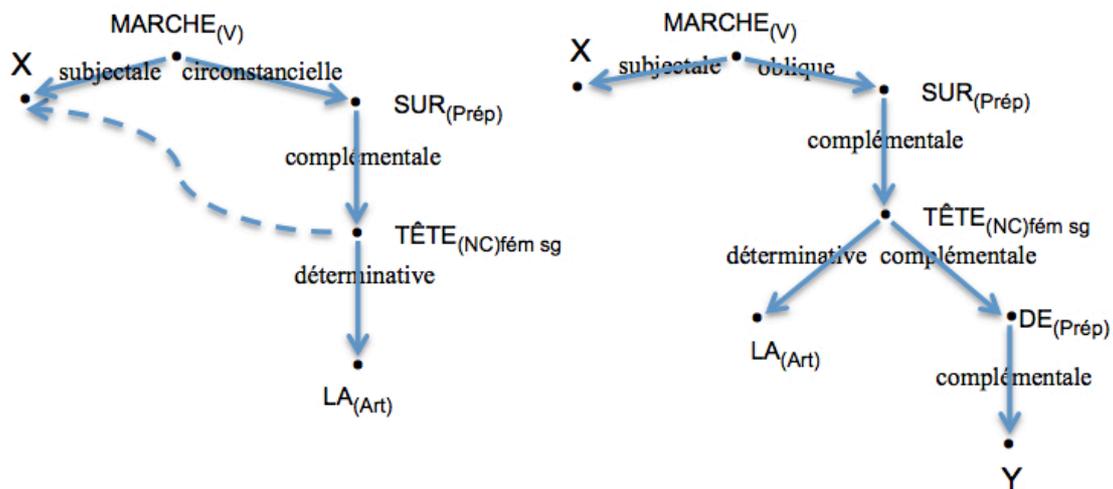


Figure 12 : Arbres de dépendance des locutions MARCHER SUR LA TÊTE ‘avoir un comportement déraisonnable’ (à gauche) et MARCHER SUR LA TÊTE ‘avoir un comportement dominant à l’égard de quelqu’un’ (à droite).

La flèche en pointillés ajoutée à l’arbre de dépendance de gauche indique que l’actant sémantique¹¹ de TÊTE correspond à l’actant de MARCHER. À l’inverse, dans l’arbre de droite, il est bien exprimé comme dépendant syntaxique de TÊTE et non pas du verbe. Les deux locutions ont alors une structure actancielle différente ; ce qui explique notamment la différence de comportement face à la montée du possesseur observée à la fin de la section 3.2.

Revenons à l’objet qui nous intéresse, à savoir la question de la polysémie et/ou de l’homonymie de la locution MARCHER SUR LA TÊTE. Du côté du signifié, il est facile – nous l’avons dit dans la section 3.2 – d’identifier la relation sémantique entre la locution de l’exemple (1) et celle de (2). Il est en revanche problématique de construire un pont sémantique reliant les deux acceptions illustrées dans les exemples (1) et (2) et MARCHER SUR LA TÊTE en (3). Ajoutons à cela des SLS différentes : ces syntagmes comprennent des constituants lexicaux et des réseaux de dépendances propres (figure 12). C’est dire que, malgré l’homophonie et l’homographie, le signifiant n’a pas la même forme. L’absence de lien sémantique de copolysémie s’accompagne de différences structurelles dans la forme du signifiant. Force est dès lors de reconnaître l’existence de deux vocables construits sur des lexies différentes du verbe MARCHER : un MARCHER SUR LA TÊTE¹, qui possède deux acceptions copolysèmes (MARCHER SUR LA TÊTE¹ I et MARCHER SUR LA TÊTE¹ II dans les exemples 1 et 2) et son homonyme MARCHER SUR LA TÊTE² (exemple 3) avec un seul sens. MARCHER SUR LA TÊTE¹ et MARCHER SUR LA TÊTE² sont morpho-phonologiquement identiques, mais lexico-syntaxiquement différentes puisque construites sur des lexies distinctes, avec des combinatoires propres¹².

Lorsque l’on conclut, comme pour MARCHER SUR LA TÊTE¹, à la polysémie, se pose naturellement la question de savoir laquelle des acceptions copolysèmes, en l’occurrence ‘infliger une lourde défaite à quelqu’un’ ou bien ‘avoir un comportement dominant à l’égard de quelqu’un’ doit être considérée comme ULB. Dans la mesure où le sens lexical décrit dans le *RL-fr* est abordé dans une perspective synchronique (cf. section 3.1 ci-dessus), il s’agit d’identifier l’acception la plus saillante pour un locuteur contemporain du français. C’est elle qui permettra ensuite de construire des ponts sémantiques entre les copolysèmes d’un vocable, et par conséquent, structurer le graphe qu’ils forment.

Nous reconnaissons volontiers qu'il n'est pas toujours facile de décider laquelle des deux acceptions de MARCHER SUR LA TÊTE¹ constitue l'ULB de ce vocable. On pourrait ainsi se tourner vers les corpus disponibles, en admettant que le sens le plus fréquent est le plus saillant pour l'ensemble des locuteurs. Le travail lexicographique développé dans le cadre du *RL-fr* exploite régulièrement trois corpus : Frantext (restreint aux ouvrages écrits à partir de 1950), le Corpus Français du Web (FrWac) et celui de l'Est Républicain (ER), sans renoncer aux exemples venant d'autres sources telles que les médias, le cinéma, le théâtre, le Web, la presse et diverses lectures personnelles. L'exploration de ces ressources a pour objectif de ramener des illustrations d'emplois.

Face à une pluralité de sens qu'il s'agit d'organiser en un réseau de copolysèmes, on pourrait utiliser des données fréquentielles, en admettant que l'ULB correspondra à l'emploi le plus fréquent de la locution. Le recours à la fréquence d'attestation en corpus présente cependant de nombreuses difficultés. Notons qu'aucun des corpus utilisés n'est désambiguïsé, ce qui implique un dépouillement manuel des résultats. Quoiqu'abordable pour une séquence telle que MARCHER SUR LA TÊTE, appliquée à l'échelle d'une ressource à large couverture, cette démarche ralentirait sérieusement l'encodage des données lexicales, sans garantir une réponse certaine à notre question. Aussi, la séquence *marcher sur la tête* est-elle attestée 44 fois dans nos trois corpus de référence, respectivement 13 fois dans Frantext, 16 fois dans le corpus de l'ER, 15 fois dans le FrWac. Il nous a fallu cependant écarter onze occurrences, soit parce qu'elles se sont avérées de nature collocationnelle (référant par exemple à des comportements peu louables de footballeurs à l'égard des membres de l'équipe adverse), soit parce qu'elles correspondaient à des doublons fréquents dans le FrWac. Nous avons également éliminé une métaphore non lexicalisée, créée *ad hoc*, et deux emplois entre guillemets. Parmi les 33 occurrences retenues, une majorité (26) instancie le sens de 'avoir un comportement déraisonnable' du vocable monosémique MARCHER SUR LA TÊTE². Les données relatives aux deux sens copolysèmes de MARCHER SUR LA TÊTE¹ sont ainsi trop peu nombreuses pour servir de base à un raisonnement.

Plutôt que de recourir à la fréquence d'emploi d'un sens, le lexicographe dépouille donc l'organisation de la polysémie sur le plan sémantique. En effet, les deux sens de MARCHER SUR LA TÊTE¹ sont reliés par un lien métaphorique dont l'ULB constitue la source. Cette métaphore s'explique, si l'on considère comme source l'acception 'infliger une lourde défaite à quelqu'un'. Son copolysème MARCHER SUR LA TÊTE¹ II se laisse ainsi paraphraser 'avoir un comportement dominant à l'égard de quelqu'un, comme si on lui infligeait une lourde défaite'. Il serait, au contraire impossible d'explicitier le fonctionnement de l'extension polysémique à partir de 'avoir à l'égard de quelqu'un un comportement'.

Notons au passage qu'un vocable lexémique, tel que DOMINER, contient des acceptions quasi-synonymes de ces locutions, employées dans les exemples (11) et (12). L'encodage systématique de liens de copolysémie dans le *RL-fr* permet de comparer leur organisation sémantique, même si celle-ci devra faire objet d'une autre étude.

11. Les leçons des révolutions du XVIII^e siècle, les déclarations des droits de l'homme, les États européens les ont enseignées dans les écoles qu'ils ont ouvertes au profit des enfants des pays qu'ils **dominaient**. (Frantext)
12. Énorme coup de bambou pour l'équipe de France qui **a été** complètement **dominée** individuellement et collectivement au Stade de France pour le match d'ouverture de « Sa » Coupe du Monde, par une équipe d'Argentine impressionnante de maîtrise (17-12). (FrWac)

5 Conclusions

Nous avons présenté ici la modélisation particulière de la polysémie dans le *RL-fr*, plus précisément celle des locutions. Nous avons montré à travers le cas de MARCHER SUR LA TÊTE comment la structure lexico-syntaxique des locutions peut aider à trancher entre polysémie et homonymie. Nous sommes par conséquent amenés à envisager une conception élargie du signifiant d'une locution. En tant que syntagme, la locution est constituée de lexies. Son signifiant ne se limite alors pas au versant morpho-phonologique (face graphique et sonore du signe saussurien), mais s'analyse à un niveau lexico-syntaxique : deux locutions peuvent se prononcer et s'écrire de la même façon sans être complètement identiques d'un point de vue formel. Elles peuvent en effet être construites sur des lexies différentes appartenant au même vocable, comme c'est le cas de MARCHER SUR LA TÊTE. Elles sont alors homographes et homophones, sans pour autant être lexicalement identiques (cf. la notion de *quasi-signe* chez Kahane, 2015 : 45, pour désigner notamment les constituants lexicaux des locutions). Ceci nous permet d'avoir un critère additionnel pour trancher de façon sûre entre polysémie et homonymie ; critère qui n'a pas cours pour les lexèmes.

Il existe toutefois des cas, rares à notre connaissance, où des locutions sont entièrement identiques — graphiquement, phonologiquement et lexicalement — sans pour autant avoir un lien sémantique identifiable. C'est le cas de LOUP DE MER, qui a au moins trois acceptions¹³ : 1) 'marin endurci et expérimenté', 2) 'espèce de poisson' et, par métonymie, 3) 'chaire de ce poisson, que l'on consomme'. S'il est aisé de lier sémantiquement les sens 2 et 3, il est moins facile de les rattacher au sens 1, qui doit alors être considéré comme appartenant à un autre vocable.

Il est fort à parier que l'exploitation des structures lexico-syntaxiques réserve aux lexicographes plus d'une surprise !

Références bibliographiques

- Bosredon, B. (2004). Le paramètre catégoriel dans les unités polylexicales : de la polysémie à la néologie, *Syntaxe et sémantique*, 1 (5), 47–58
- Collectif (2004). *Syntaxe et sémantique*, 1(5). Caen : Presses Universitaires de Caen.
- Dubois, J., Dubois-Charlier, F. (1997). *Les Verbes français*. Paris : Larousse-Bordas.
- Fellbaum, C. (éd.) (1998). *WordNet : An Electronic Lexical Database*. Cambridge (MA) : MIT Press.
- Fillmore, C., Johnson, C., Petruck, M. (2003). Background to Framenet, *International Journal of Lexicography*, 16 (3), 235–250.
- Gader, N., Lux-Pogodalla, V., Polguère, A. (2012). Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor. *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex III), The COLING 2012 Organizing Committee, Mumbai*, 109–125.
- Gaume, B. (2004). Balades aléatoires dans les petits mondes lexicaux. *Information interaction intelligence*, 4(2), 39–96.
- Kahane S. (2015). Les trois dimensions d'une modélisation formelle de la langue : syntagmatique, paradigmatique et sémiotique, *TAL*, 56 (1), 39–63.
- Kiss, G. R. (1968). Words, associations, and networks. *Journal of Verbal Learning and Verbal Behavior*, 7, 707-713.
- Lux-Pogodalla, V. Polguère, A. (2011). Construction of a French Lexical Network. Methodological Issues. *Proceedings of the First International Workshop on Lexical Resources, WoLeR 2011. An ESSLLI 2011 Workshop, Ljubljana*, 54–61.
- Mel'čuk, I. (2004a). Actants in Semantics and Syntax I : Actants in Semantics, *Linguistics*, 42 (1), 1–166.
- Mel'čuk, I. (2004b). Actants in Semantics and Syntax II : Actants in Syntax, *Linguistics*, 42 (2), 247–291.
- Mel'čuk, I. (1996). Lexical Functions : A Tool for the Description of Lexical Relations in the Lexicon. In : Wanner, L. (éd.) *Lexical Functions in Lexicography and Natural Language Processing, Language Companion Series*, 31, Amsterdam/Philadelphia : John Benjamins, 37–102.
- Mel'čuk, I., Clas, A., Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.
- Ollinger, S. (2014). Regroupement de structures de dérivations lexicales par raisonnement analogique. In : Bigi, B. (éd.) *Actes de Récital 2014, Faculté Saint Charles, Aix Marseille Université (Marseille), du 1er au 4 juillet 2014*, 92–103.
- Osherson, A., Fellbaum, C. (2009). The representation of idioms in WordNet. *Actes de ACL-IJCNLP 09*.
- Pausé, M.-S. (à paraître). Modélisation de la structure lexico-syntaxique des locutions au sein d'un réseau lexical. In : Kauffer, M., Keromnes, Y. (éd.) *Approches théoriques et empiriques en phraséologie. Actes du colloque international (Nancy, 11-12 décembre 2014)*. Tübingen : Stauffenburg.
- Pausé, M.-S. (en préparation). *Modélisation de la structure lexico-syntaxique des locutions françaises au sein d'un réseau lexical*. Thèse, Université de Lorraine, ATILF-CNRS (UMR 7118), Nancy.
- Perrin, L. (2014). Citation, lexicalisation et interprétation des expressions idiomatiques. In : Authier-Revuz, J., Doury, M. (éd.) *Parler des mots : le fait autonymique en discours*. Paris : Presse Sorbonne Nouvelle, 281–290.
- Polguère, A. (1997). Meaning-Text Semantic Networks as a Formal Language. In : Wanner, L. (éd.) *Recent Trends in Meaning-Text Theory*. Amsterdam, Philadelphie : John Benjamins, 1–24.
- Polguère, A. (2009). Lexical systems: graph models of natural language lexicons. *Language Ressources and Evaluation*, 43, 41–55.
- Polguère, A. (2011). Classification sémantique des lexies fondée sur le paraphrasage. *Cahiers de Lexicologie*, 98 (1), 197–211.

- Polguère, A. (2014). From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography*, 27 (4), 396–418.
- Pruvost, J. (2011).) À travers siècles, un article de dictionnaire et ses séquences figées. In : Anscombre, J.-C., Mejri, S. (éd.) *Le figement linguistique : la parole entravée*. Paris : Honoré Champion, 341-362.
- Selva, T., Verlinde, S., Binon, J. (2003). Vers une deuxième génération de dictionnaires électroniques. *Traitement Automatique des Langues*, 44 (2), 177–197.
- Sikora, D. (2014). Brève histoire d'une erreur lexicale. Polysémie et liens lexicaux dans l'enseignement du vocabulaire. In : Neveu, F., Blumenthal, P., Hriba, L., Gerstenberg, A., Meinschafer, J., Prévost, S. (éd.) *4e Congrès Mondial de Linguistique Française – CMLF 2014*. Paris : Institut de Linguistique Française, 1157–1172.
- Wilkins, D. (1996). Natural Tendencies of Semantic Change and the Search for Cognates. In : Durie, M., Ross, M. (éd.) *The Comparative Methods Reviewed. Regularity and Irregularity of Semantic Change*. New York / Oxford : Oxford University Press, 264–304.

Ressources en ligne :

Frantext : <http://www.frantext.fr/>

FrWac : <http://nl.ijs.si/noske/index.html>

Est Républicain : <http://www.cnrtl.fr/corpus/estrepublikain/>

¹ Nous remercions Sandrine Ollinger pour les précieuses statistiques de la base lexicale qu'elle nous a fournies, sur lesquelles nous nous sommes appuyées pour ce travail. Nous remercions également Alain Polguère pour les discussions fructueuses que nous avons eues ensemble, ainsi que Véronika Lux-Pogodalla pour ses retours sur une première version de cet article. Nos remerciements vont aussi aux relecteurs anonymes : leurs remarques nous ont permis d'améliorer le texte. Nous sommes entièrement responsables de toute erreur pouvant subsister.

² Le Wiktionnaire, ressource collective enrichie et modifiée par la communauté internet, se distingue des grands dictionnaires tels que le Robert et le Larousse, en accordant aux locutions leurs propres articles. (<http://fr.wiktionary.org>)

³ TELEPHONE PORTABLE est identifiée comme locution dans Antidote et Wiktionnaire. Aucune catégorie grammaticale n'est spécifiée dans le Robert. Néanmoins, ce dernier considère POMME DE TERRE, unité polylexicale également analogue à un syntagme, comme un nom commun. Quant au Larousse, dans sa version en ligne il n'offre aucune correspondance pour cette unité. Pour une description morphologique – et donc une conception comme nom composé – voir Bosredon (2004).

⁴ Pour un descriptif succinct, voir : <http://www.atilf.fr/spip.php?article908>

⁵ Un système lexical de l'anglais est également en cours d'élaboration. Bien que dans un état de développement moins avancé, des bases du coréen et de l'espagnol ont été développées.

⁶ Pour une présentation de la notion d'*étiquette sémantique*, voir Polguère (2011). Rappelons simplement qu'il s'agit d'une caractérisation minimale classifiante du sens décrit : les lexies telles que AUTO, AUTOCAR, BUS, CAMION, TRAMWAY, VOITURE, etc. portent l'étiquette 'véhicule'.

⁷ Le travail d'élaboration des définitions est actuellement dans une phase initiale. Dans la mesure où il s'agit d'une démarche strictement normalisée et contrôlée, il est en effet indispensable de mettre en place une procédure de rédaction explicite et une feuille de route précise.

⁸ Nous empruntons cette paraphrase définitoire au *Petit Robert 2011* dans sa version électronique.

⁹ Notons que ce score est encore bien inférieur à celui de la base *Les verbes français* de Dubois et Dubois-Charlier (1997), qui décrit 60 acceptions du verbe PASSER.

¹⁰ Notons que la polysémie de certaines lexies dites grammaticales n'est à ce jour pas suffisamment travaillée. C'est la raison pour laquelle, dans certaines SLS, nous ne pointons pas encore vers la lexie, mais affichons uniquement la forme.

¹¹ Sur l'opposition entre actants sémantiques et syntaxiques, voir notamment Mel'čuk (2004a et 2004b)

¹² MARCHER SUR LA TÊTE n'est, bien entendu, pas un cas isolé. Nous pouvons citer, entre autres, l'exemple de COUP DE BARRE. La locution a au moins deux sens : 'fatigue subite' (le fameux coup de barre provoqué par la digestion post-méridionale) et 'changement' (le coup de barre que l'on donne à sa vie en changeant de profession). Si le second sens est très nettement construit par analogie avec le changement de cap d'un bateau à l'aide de la barre actionnant le gouvernail, le premier est vraisemblablement construit par analogie avec un coup de barre (de fer, par exemple) assommant, reçu sur la tête. Les SLS seront alors différentes et nous aurons deux vocables homonymes.

¹³ Au moins, car les encyclopédies utilisent ce syntagme comme nom vernaculaire de plusieurs espèces, qu'il s'agisse de phoques ou de poissons.