

# Modélisation des locutions dans un réseau lexical : enjeux théoriques

Marie-Sophie PAUSÉ  
ATILF-CNRS

## *Résumé*

Cet article présente les impacts théoriques de la description lexicographique des unités phraséologiques du français – en particulier les locutions – au sein d’un réseau lexical construit suivant les principes de la Lexicologie Explicative et Combinatoire. La description dans un tel cadre théorique prend en compte les relations syntagmatiques et paradigmatices entre les unités lexicales – lexèmes et locutions – dans l’objectif de rendre compte de la façon dont elles se combinent. Après un panorama des différents types d’unités phraséologiques nous en présentons les principales caractéristiques de description. Nous nous attardons sur le cas des locutions dont les spécificités de la structure lexicosyntaxique doivent être prises en considération. Ceci implique l’identification d’un statut lexical, d’unités lexicales constituantes et d’une structure syntaxique.

**Mots clés :** lexicologie, phraséologie, locutions, combinatoire.

## *Abstract*

This paper deals with the theoretical impacts of lexicographical description of French phraseological units – idioms in particular – within a lexical network constructed according to the principles of Explanatory and Combinatorial Lexicology. Such a description takes into account syntagmatic and paradigmatic links between lexical units – lexemes and idioms – in order to explain how they combine. After an overview of the different types of phraseological units, we present the main characteristics of their description. We focus on the case of idioms whose specific lexico-syntactic structures must be considered. This implies identification of a lexical status, a syntactic structure and lexical constituents.

**Keywords:** lexicology, phraseology, idioms, combinatory.

## **Resumen**

En este trabajo se presentan los impactos teóricos de la descripción lexicográfica de las unidades fraseológicas del francés –en particular las locuciones– dentro de un recurso léxico construido según los principios de la Lexicología Explicativa y Combinatoria. La descripción en tal marco teórico tiene en cuenta las relaciones sintagmáticas y paradigmáticas entre las unidades léxicas –lexemas y locuciones– con el objetivo de dar cuenta de cómo se combinan. Después de una visión general de los diferentes tipos de unidades fraseológicas, presentamos las principales características de sus descripciones. Nos centramos en el caso de las locuciones cuyas especificidades relativas a la estructura léxico-sintáctica deben tenerse en cuenta. Esto implica la

identificación de un estatus léxico, de las unidades léxicas constitutivas y de una estructura sintáctica.

**Palabras clave:** lexicología, fraseología, expresiones idiomáticas, combinatoria.

## Introduction

La phraséologie est devenue une des préoccupations importantes des lexicologues et lexicographes (Blanco, 2014 ; Pruvost, 2011 ; Osherson et Fellbaum, 2010 ; Fellbaum et Geyken, 2005). La récurrence et la variété des phénomènes phraséologiques (González Rey, 2015 ; Mel'čuk, 2013) doivent être prises en compte dans des ressources lexicales de grande envergure.

Nous proposons dans notre contribution de présenter les impacts théoriques de la description lexicographique des locutions au sein d'un réseau lexical construit suivant les principes de la Lexicologie Explicative et Combinatoire (Polguère, 2014 ; Mel'čuk, Clas, et Polguère, 1995). La description dans un tel cadre théorique prend en compte les relations syntagmatiques et paradigmatisées entre les unités lexicales – lexèmes et locutions – dans l'objectif de rendre compte de la façon dont elles se combinent.

Conventions d'écriture : les unités lexicales – lexies – apparaîtront en petites majuscules – SOURIS – et seront pourvues de leur numérotation lexicographique issue de la ressource dont nos données sont issues (le Réseau lexical du français) – SOURIS I. Le signifié apparaîtra entre guillemets simples 'petit mammifère rongeur...'. Les locutions seront indiquées par des crochets : 'DONNER SA LANGUE AU CHAT'.

## Des unités phraséologiques

La variabilité des phénomènes phraséologiques produit des unités ayant des caractéristiques de fonctionnement différentes qu'il est nécessaire de prendre en compte dans une ressource lexicale. En effet, certains syntagmes phraséologiques sont sélectionnés comme un tout pour signifier un message précis. C'est le cas de 'METTRE LA CHARRUE AVANT LES BŒUFS' qui signifie 'anticiper quelque chose qui n'est pas encore envisageable' et *un train peut en cacher un autre* qui, affiché à un passage à niveau, signifie que deux trains peuvent passer en même temps. D'autres syntagmes phraséologiques sont issus de la combinaison entre une unité

choisie librement, et une autre unité choisie en fonction de la première pour véhiculer un sens précis. Par exemple, on associe le verbe TOMBER V.1a à NUIT I, mais on ne peut pas l'associer à JOUR I : \**le jour tombe*.

La (non-)compositionnalité sémantique permet de distinguer les syntagmes comme 'METTRE LA CHARRUE AVANT LES BŒUFS' des syntagmes comme *un train peut en cacher un autre* et *la nuit tombe*. En effet, le sens de 'METTRE LA CHARRUE AVANT LES BŒUFS' n'est pas la somme stricte des sens de ses composants, tandis que c'est le cas pour les deux autres syntagmes.

Qu'ils soient sémantiquement compositionnels ou non, nous appellerons désormais les syntagmes phraséologiques *phrasèmes*, définis comme des énoncés multilexémiques non libres (Mel'čuk, 2013). Les phrasèmes sémantiquement non compositionnels sont appelés *locutions*, les phrasèmes sémantiquement compositionnels semi-contraints comme *la nuit tombe* sont appelés *collocations*<sup>1</sup>, et les phrasèmes analogues à *un train peut en cacher un autre* sont nommés *clichés linguistiques*.

Nous nous intéressons ici à la modélisation lexicographique des phrasèmes, en particulier celle des locutions.

## Phraséologie et lexicographie

La prise en compte des phrasèmes dans les ressources lexicales ne cesse de croître. Des ressources comme WordNet (Fellbaum, 1998) et FrameNet (Ruppenhofer et al., 2010) ont pleinement intégré les locutions en tant qu'entrées à part entière de leurs nomenclatures. Nous regrettons toutefois le statut marginal de ces unités dans les ressources lexicales de référence pour le français, à savoir les dictionnaires généraux tels que Larousse et Le Robert et le Trésor de la Langue Française. Les différentes catégories de phrasèmes ne sont pas toujours bien délimitées et leurs spécificités ne sont donc pas prises en compte.

## Ce qui a été fait

WordNet est un type de système lexical qui propose une description fine de la synonymie entre les unités lexicales – lexèmes et locutions

---

<sup>1</sup> Notons qu'il existe une autre conception de la collocation issue de l'approche statistique de Firth (1957), utilisée entre autres en linguistique de corpus (Sinclair, 1981). Pour une comparaison des deux approches, on pourra consulter Evert (2005).

(Fellbaum, 1998 ; Osherson et Fellbaum, 2010). Les unités synonymes sont regroupées au sein d'un *synset* et classées du sens le plus neutre au sens le plus stylistiquement marqué. Les unités de différents *synsets* sont liées entre elles par des liens morpho-syntaxiques et sémantiques. La ressource a d'abord été développée pour l'anglais, avant d'être étendue à d'autres langues.

FrameNet est une ressource basée sur les principes des grammaires de construction et des « *frame semantics* » (lit. « cadres sémantiques ») (Fillmore, Kay, et O'Connor, 1988 ; Fillmore, 2008). Elle contient une description des propriétés sémantiques, syntaxiques et combinatoires des unités lexicales. Chaque unité lexicale – lexème ou locution – se voit décrite en fonction des constructions dans lesquelles elle se réalise. Les cadres permettent de décrire des situations ainsi que les participants qu'elles impliquent. Par exemple, 'VENDRE LA MÈCHE' implique trois participants : un locuteur, un interlocuteur, une information révélée. On peut également ajouter le canal par lequel le locuteur révèle l'information (par téléphone, par lettre, etc.) et le sujet sur lequel porte l'information. Ces participants sont appelés *frame elements* (lit. « éléments du cadre »). Tout comme WordNet, FrameNet fut initialement consacrée à la description de l'anglais, puis étendue à d'autres langues<sup>2</sup>.

Pour le français, les travaux effectués dans le cadre du Lexique-Grammaire (Gross, 1975) ont mis à la disposition des linguistes une série de patrons syntaxiques d'unités phraséologiques (Tolone, 2011).

Les ressources ici énumérées ont en commun le fait de donner une place importante aux phrasèmes, particulièrement aux locutions. Néanmoins, elles ne donnent qu'une description partielle de leur combinatoire : une description plutôt sémantique en ce qui concerne WordNet et FrameNet, et une description syntaxique en ce qui concerne le Lexique-Grammaire. Nous verrons par la suite que pour donner toutes les clés d'utilisation des unités phraséologiques aux locuteurs ou apprenants d'une langue, il est nécessaire d'adopter une description à l'interface entre sémantique et syntaxe.

## Réseau lexical du français

Le Réseau lexical du français (Lux-Pogodalla et Polguère, 2011 ; Gader, Lux-Pogodalla, et Polguère, 2012 ; Polguère, 2014), désormais RL-fr, est une

<sup>2</sup> [https://framenet.icsi.berkeley.edu/fndrupal/framenets\\_in\\_other\\_languages](https://framenet.icsi.berkeley.edu/fndrupal/framenets_in_other_languages)

ressource lexicale développée au laboratoire d'*Analyse et Traitement Informatique de la Langue Française*<sup>3</sup> suivant les principes de la Lexicologie Explicative et Combinatoire (Mel'čuk, Clas, et Polguère, 1995)<sup>4</sup>. Les phrasèmes sont sous-catégorisés et décrits en fonction de leurs caractéristiques propres.

## Phrasèmes dans le Réseau Lexical du Français

Il est nécessaire dans un premier temps de distinguer différents types d'unités phraséologiques dans le souci de rendre compte de leurs particularités. Toutes les unités phraséologiques n'ont pas besoin d'une définition lexicographique, par exemple. Nous nous proposons d'introduire la classification des phrasèmes adoptée dans le cadre de la constitution du RL-fr.

### Nœuds phraséologiques

Les nœuds phraséologiques du RL-fr sont à subdiviser en deux types :

1. les nœuds qui sont des unités lexicales
2. les nœuds qui ne sont pas des unités lexicales.

La distinction entre ces deux types de nœuds est basée sur un seul critère : la (non-)compositionnalité sémantique. Tout syntagme dont la combinatoire est restreinte et dont le sens est non compositionnel est considéré comme une locution – BALAYER DEVANT SA PORTE 'considérer ses propres manquements avant de regarder ceux des autres', OR NOIR 'pétrole', PETIT A PETIT 'progressivement'. La locution est une unité lexicale en tant qu'elle a des caractéristiques grammaticales et une définition. Dans le RL-fr, le proverbe est un type particulier de locution.

Un syntagme dont la combinatoire est restreinte et dont le sens est compositionnel peut être :

1. un cliché linguistique
2. une collocation

Un cliché linguistique est un syntagme sémantiquement compositionnel et formellement contraint qui peut constituer à lui seul

<sup>3</sup> ATILF, UMR, Nancy, France.

<sup>4</sup> L'état actuel de la nomenclature est disponible sur la plateforme Ortolang.

un énoncé : *L'abus d'alcool est dangereux pour la santé*. Chaque cliché linguistique est considéré comme un nœud du RL-fr.

Une collocation est, selon Polguère (2016, 65), « *un syntagme AB (ou BA) qui est tel que, pour le construire, le Locuteur sélectionne A librement d'après son sens 'A', alors qu'il sélectionne B pour exprimer auprès de A un sens 's' en fonction de contraintes imposées par A.* », comme *peur bleue, rendre un jugement, la nuit tombe*. Les collocations sont encodées comme des liens entre des nœuds du réseau au moyen des fonctions lexicales syntagmatiques Sens-Texte. Les fonctions lexicales permettent de modéliser les liens paradigmatiques et syntagmatiques entre les lexies. Faute de place, nous en ferons une présentation très sommaire. Le lecteur intéressé pourra consulter Mel'čuk (2007). Le tableau 1 contient quelques exemples de liens avec les fonctions lexicales qui permettent de les encoder.

Liens		Exemples de liens entre lexies	Fonctions lexicales
~ paradigmatiques	synonymie	voiture → <b>fam</b> bagnole	<b>Syn</b> (voiture) = <b>abrégé</b> <i>auto, automobile, fam bagnole, fam caisse</i>
	nom du premier actant	voiture → conducteur, conductrice	<b>S<sub>1</sub></b> (voiture) = <i>conducteur, conductrice</i>
~ syntagmatiques	verbe d'utilisation	voiture → conduire <sub>[Samuel conduit sa voiture]</sub>	<b>Real<sub>1</sub></b> (voiture) = <i>conduire</i> [ART ~], <i>rouler</i> [en ~ ]/[dans ART ~ ], <i>voyager</i> [en ~ ]/[dans ART ~ ]
	verbe de fonctionnement	voiture → rouler <sub>[La voiture roule]</sub>	<b>Facto</b> (voiture) = <i>avancer, rouler</i>

Tableau 1: Exemples d'utilisation des fonctions lexicales

Les fonctions lexicales simples peuvent être combinées entre elles pour obtenir des fonctions lexicales complexes. Par exemple, **LiquFacto** (*page*) = *tourner* indique que *tourner* est le verbe utilisé pour signifier 'rendre la page non fonctionnelle'.

Notre tâche de lexicographe phraséologue nous a amenée à envisager une quatrième classe de phrasèmes. Notre objectif à long terme étant de prendre en compte les caractéristiques combinatoires

inhérentes à chaque type de phrasème, nous nous sommes appuyée sur des données attestées. Partant d'une liste de plus de 3.000 candidats au statut de *locution*, nous avons recherché les emplois de plusieurs échantillons de cette liste dans des bases textuelles telles que Frantext<sup>5</sup>, l'Est Républicain et FrWac (Baroni et al., 2009) ainsi que plus largement sur le web à l'aide d'un moteur de recherche. Nous avons pu alors distinguer deux types de variantes paradigmatiques :

1. variantes avec un paradigme restreint
2. variantes avec un paradigme ouvert.

Le premier cas de figure peut être illustré par 'TAILLER UNE VESTE' 'critiquer pour nuire à la réputation'. Le constituant *veste* peut en effet y être remplacé par d'autres noms appartenant à la classe des vêtements, comme l'illustrent les exemples :

- (1) a. Après s'être fait **tailler une veste** chez Jean-Marc Morandini, le jeune homme s'est fait remettre à sa place par Sinik, l'un des maîtres incontestés du rap en France. [FrWac]
- b. Je trouve cela assez injuste mais comme je ne veux pas leur faire le plaisir de disparaître, je continue d'accepter les invitations télé et ce même si c'est pour me faire **tailler un costume**. [FrWac]
- c. Domenech **taille un short** à Wenger. [FrWac]

À notre connaissance, seules ces trois variantes sont possibles. Elles n'altèrent pas le sens de la locution mais permettent certains défigements. Par exemple, lorsque la locution est employée avec *costume* en tant que complément du verbe *tailler*, le nom peut être complété par des unités compatibles avec lui, comme le montrent les occurrences :

- (2) a. Face à ses nombreux thuriféraires qui attendaient la scène, le président de l'UMP Nicolas Sarkozy s'est fait **tailler un costume trois pièces**, en bonne et due forme. [FrWac]
- b. François s'est longtemps plaint de la désaffection des journalistes. Maintenant il devrait les bénir de lui **tailler un costume sur mesure**. [FrWac]

---

<sup>5</sup> [www.frantext.fr/](http://www.frantext.fr/)

Nous considérons que *tailler une veste*, *tailler un costume* et *tailler un short* sont trois variantes d'une même locution pouvant être notée «TAILLER UNE {VESTE}»<sup>6</sup>.

Le second type de variantes paradigmatiques est illustré par les exemples :

- (3) a. Une piscine couverte par un abri coûte **dans les** 35.000 €. [Web, [www.guide-piscine.fr/prix-piscine-enterree/combien-coute-une-piscine-enterree-ou-creusee-1173\\_A](http://www.guide-piscine.fr/prix-piscine-enterree/combien-coute-une-piscine-enterree-ou-creusee-1173_A), consulté le 12/04/2020]
- b. Les plus larges envergures au monde sont celles de zygoptères sud et centre américains atteignant **dans les 30 cm**.  
[[www.insecte.org/forum/viewtopic.php?t=105405](http://www.insecte.org/forum/viewtopic.php?t=105405), consulté le 12/04/2020]
- (4) a. Son sourire d'ailleurs **en disait long** sur l'état de ses relations avec la vaincue. [Frantext ; Bazin Hervé, *Cri de la chouette*, 1972, 36]
- b. On l'a foutu au rancart voilà deux ans. Il rêve d'en sortir par la grande porte et, comme par hasard, vient d'apprendre qu'un journaliste suisse, qui **en sait long** sur un trafic d'objets d'art, rôde dans Paris à propos de la mort de Langlet. [Frantext ; Embareck Michel, *Sur la ligne blanche*, 1984, 163]
- (5) Si je cherchais à corrompre le commandant, ce ne serait pas avec des bœufs et des œufs, il y a mieux pour cela ; mais il est de coutume que le passage d'un chef soit pour les pauvres l'occasion de manger **de la** viande et de boire **du** lait à satiété. [Frantext ; Bâ Amadou Hampâté, *Oui mon commandant*, 1994, 246]

*dans les* + SN, employé dans les énoncés constitue un bel exemple de non correspondance entre niveau sémantique et niveau syntaxique. La combinaison forme un syntagme prépositionnel introduit par *dans*. Sémantiquement, dans ces énoncés, le complément nominal de la préposition est l'actant du verbe dont le syntagme prépositionnel est syntaxiquement dépendant. Un schéma vaudra mieux que des paroles. La figure 1 montre (à gauche) une partie de la représentation sémantique de *les zygoptères atteignent dans les 30 cm* (cf. exemple 3b), et (à droite) la structure syntaxique de l'énoncé.

<sup>6</sup> Le lecteur intéressé pourra trouver une analyse plus approfondie de cette locution, associée au mécanisme de séparation phraséologique dans Pausé et Polguère (à paraître).

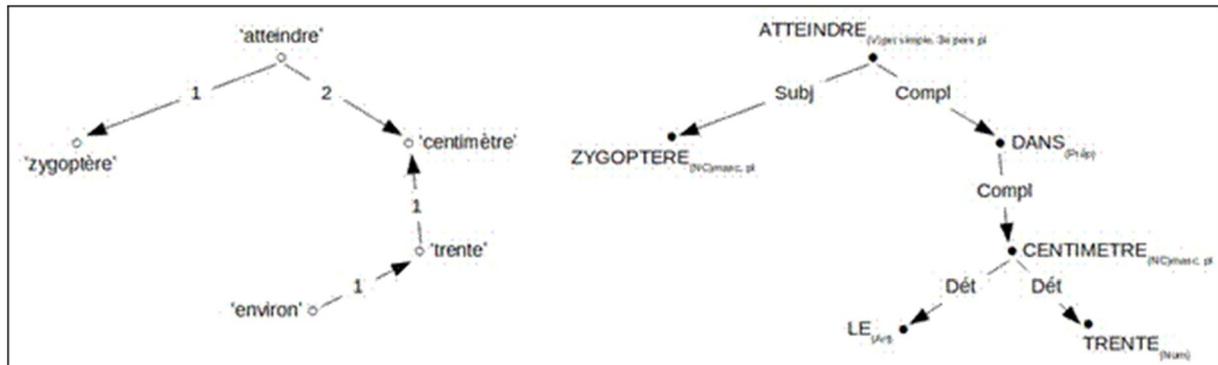


Figure 1 : Extraits des représentations sémantique et syntaxique de « Les zygoptères atteignent dans les trente centimètres »

Le lexicographe se retrouve donc avec une unité significative syntagmatique dont le noyau syntaxique est un paradigme ouvert. Bien entendu, n'importe quelle lexie ne peut pas occuper la place libre. La lexie doit exprimer un paramètre de mesure pouvant être approximatif. Les deux unités fixes (*dans* et *les*), dont la première est le gouverneur syntaxique de la lexie libre et la seconde son dépendant syntaxique, permettent d'exprimer auprès de la lexie en question l'approximation. Le sens de *dans les* + *SN* n'est pas la somme de ses constituants : la combinaison est sémantiquement non compositionnelle.

Pausé (2017, 96) propose une nouvelle catégorie de phrasèmes prenant en compte les combinaisons de ce type : le phrasème non connexe, défini comme suit :

*Un phrasème non connexe est un phrasème dont le signifiant est un syntagme incomplet, dont certaines positions lexicales sont remplies par des unités lexicales fixes – principalement des lexies grammaticales – alors que d'autres sont ouvertes à un paradigme plus ou moins restreint. Les positions ouvertes peuvent occuper la place de la tête de syntagme, ou bien celle d'un dépendant syntaxique. Les combinaisons peuvent être sémantiquement compositionnelles, ou non.*

Cette catégorie de phrasème est fortement inspirée de la notion de *construction* au sens large, telle qu'envisagée dans les grammaires de construction (Fillmore, Kay, et O'Connor, 1988). Les constructions au sens large incluent les constructions dites *grammaticales* dont la théorisation remonte au moins au XVIII<sup>e</sup> siècle. Diderot et d'Alembert (1751) définissent la construction grammaticale comme « [...] l'arrangement des mots

*dans le discours* »<sup>7</sup>. Les constructions au sens large englobent des phénomènes plus hétérogènes, parmi lesquels :

1. des verbes présentant des exigences contextuelles hors de leur projection – à savoir la réalisation des arguments ; par exemple les verbes qui requièrent des contextes impliquant à la fois capacité et négation : *can't stand*, *can't help*, etc. ;
2. des schémas en miroir, du type  $A : B = C : D$ ; *blood is to red as snow is to white* ;
3. des numéraux qualifiés, comme *a mere five dollars* (lit. 'un simple cinq dollars') 'à peine cinq dollars', *an additional thirty pages* (lit. 'une supplémentaire trente pages') 'une trentaine de pages supplémentaires'.

Nous ne pouvons hélas pas nous étendre davantage sur la notion de construction, dont les principes sont repris et commentés par Croft et Cruse (2004).

Les phrasèmes non connexes nous permettent de prendre en considération des unités phraséologiques dont le fonctionnement diffère de celui du syntagme standard. Il peut s'agir de :

1. constructions syntaxiques, comme le partitif formé de la préposition DE et de l'article LE et fonctionnant comme déterminant d'un nom.
2. constructions lexicalisées, comme *dans les* + SN, quasi-synonyme de l'adverbe d'approximation ENVIRON, ou bien EN V LONG (*en savoir long*, *en connaître long*, cf. exemples ), qui permet en synchronie à LONG de signifier 'beaucoup', uniquement en association avec EN.

Le tableau 2 récapitule les différents phrasèmes décrits dans le RL-fr, ainsi que leur statut.

---

<sup>7</sup>On parle ensuite, en syntaxe, de constructions clivées, pseudo-clivées, disloquées, etc.

Phrasèmes	Nœuds du réseau	Unités lexicales	Liens entre nœuds du réseau
Locutions	+	+	-
Clichés linguistiques	+	-	-
Collocations	-	-	+
Phrasème non connexe	+	(+)	-

Tableau 2: Récapitulatif des types de phrasèmes décrits dans le RL-fr.

Pour tenir compte des caractéristiques des différents phrasèmes, nous avons été amenée à élaborer divers statuts lexicaux.

### Statuts lexicaux

Le statut lexical d'une unité correspond à son statut au sein de l'entité « lexique ». Trois types de statuts lexicaux ont été identifiés :

1. statut phraséologique
2. statut pragmatique
3. statut langagier

Le statut phraséologique est attribué à tout nœud syntagmatique du réseau. Pour les locutions, il permet de préciser la nature de leur sens lexical. Dans une conception de la non-compositionnalité du point de vue de l'encodage d'un énoncé, le sens d'une locution peut comprendre le sens d'un, plusieurs ou aucun de ses constituants lexicaux (Mel'čuk 2013 ; Polguère 2015) :

- semi-locution : 'DÉFORMATION PROFESSIONNELLE' 'attitude acquise au cours de la vie professionnelle qui se manifeste également dans la vie privée' ;
- locution faible : 'LEVER DE RIDEAU' I 'début d'une pièce de théâtre indiqué par le lever du rideau masquant la scène'<sup>8</sup> ;
- locution forte : 'MARIAGE DE LA CARPE ET DU LAPIN' 'association de deux entités incompatibles'.

<sup>8</sup> La locution a deux autres sens : un type de pièce de théâtre, et un fait évocateur des faits suivants. Pour une présentation de la polysémie des locutions, voir Pausé et Sikora (2016).

Chacun de ces sous-types de locutions donne lieu à un statut lexical. De même, il existe un statut lexical pour les phrasèmes qui sont des clichés linguistiques.

Le statut pragmatique est attribué aux phrasèmes dont l'emploi est restreint à certains contextes extra-linguistiques. C'est le cas des pragmatèmes et des proverbes.

Le statut langagier est utilisé pour :

1. des lexies qui appartiennent à un état antérieur de la langue, comme : BALLON<sup>1II</sup> 'fesses', qualifiées de *lexies inactives* ;
2. des unités issues de règles morphologiques de dérivation comme par exemple *galopant* 'qui galope', qualifiées de *lexies régulières*.

Le RL-fr est une ressource synchronique qui a pour vocation de décrire le français du XXI<sup>e</sup> siècle. Néanmoins, lorsque nous décrivons des phénomènes phraséologiques, nous sommes amenée à toucher des phénomènes diachroniques et à manipuler des unités du lexiques qui ne font pas partie de notre nomenclature principale. Le statut langagier nous est, dans ce cas, très précieux. Nous allons voir dans la section suivante qu'il intervient tout particulièrement pour la description des locutions.

### **Particularités de la description lexicographique des locutions**

Une attention toute particulière a été portée à la modélisation formelle des locutions, unités syntagmatiques sémantiquement non compositionnelles. Une description lexico-syntaxique qui prend en compte leur signifiant, résultant de la combinaison syntaxique d'unités lexicales, a été proposée par Pausé (2017). Cette description comprend :

1. l'identification d'une partie du discours (Mel'čuk 2006) et d'un statut lexical ;
2. l'identification du patron syntaxique sur lequel elle est formée ;
3. l'identification des constituants lexicaux de la locution.

Les étapes 2 et 3 forment ce que nous appelons une structure lexico-syntaxique.

## Structure lexico-syntaxique

Les locutions sont des syntagmes, autrement dit des unités lexicales liées par des liens syntaxiques. Preuves en sont les phénomènes de flexibilité formelle que l'on peut rencontrer, comme dans les exemples :

- (6) a. **La page qu'il a tournée hier** à Audernarde, au bout de son dernier Tour des Flandres, épreuve fétiche qu'il a tant aimée, a eu quelque chose d'une passation des pouvoirs. [Web, pressreader.com]
- b. Ethan espérait **avoir tourné une page** en quittant le service actif de la Force Mission Impossible pour un poste de formateur [...]. [FrWac]
- c. Il faut savoir faire des deuils, **tourner des pages**. Et aussi arriver à se remettre en question. [Web, parismatch.com]
- d. Il est prêt à reconnaître l'indépendance du Kosovo. Pourquoi ? Non par choix idéologique – mais pour, enfin, **tourner la page de la décennie de guerre** qui a marqué le pays. [FrWac]
- e. **La page a été tournée** sur l'année 2015 au cours de laquelle il a perdu son grand-père et dû retourner précipitamment au chevet de son père, victime d'un accident vasculaire cérébral (AVC) [Web, lapresse.ca]

Ces occurrences témoignent de l'application d'une structure communicative aux locutions, comme on le fait pour les syntagmes dits libres (Mel'čuk 2001 ; 1995).

*A speaker does not normally grab a phraseme and try on it a battery of existing syntactic transformations. What he really does is make semantic choices, that is, choices that take place at the semantic level and can lead to syntactic transformations of the phraseme, such as passivization, clefting, relativization, and the like (Mel'čuk 1995, 205).*

La mise en relation des niveaux sémantique et syntaxique des locutions a également été soulevée entre autres par Nunberg et al. (1994) et Langlotz (2006) :

*For instance, prendre le taureau par les cornes can be rendered as 'tackle a problem directly' with taureau denoting the problem and prendre meaning 'tackle.' Consequently, this idiom can be passivised because taureau has its independent metaphorical meaning in the context of the phrase. The same is true for the cleft extraction and adnominal modification [...]. The constituent taureau can be*

*postmodified by the complement des privatisations due to its metaphorical sense* (Langlotz 2006, 35).

Dans le souci de décrire à terme la combinatoire d'une locution, nous nous devons de décrire son signifiant syntagmatique.

Dans le RL-fr l'attribution d'une structure lexico-syntaxique s'effectue en deux temps, une fois la partie du discours de la locution déterminée (Mel'čuk, 2006). Tout d'abord il faut déterminer son patron syntaxique. C'est en quelque sorte la charpente sur laquelle on va construire notre structure. La structure lexico-syntaxique de 'TOURNER LA PAGE' est fondée sur le patron V Art NC, tout comme celle de 'ALLONGER LA SAUCE' ou 'MANGER LE MORCEAU'. Les principes adoptés pour la modélisation du patron sont ceux des grammaires de dépendance (Kahane, 2000).

Le patron est directement lié aux unités lexicales constitutives de la locution que l'on va identifier dans un second temps. La figure 2 illustre l'attribution d'une structure lexico-syntaxique à 'ALLONGER LA SAUCE'.

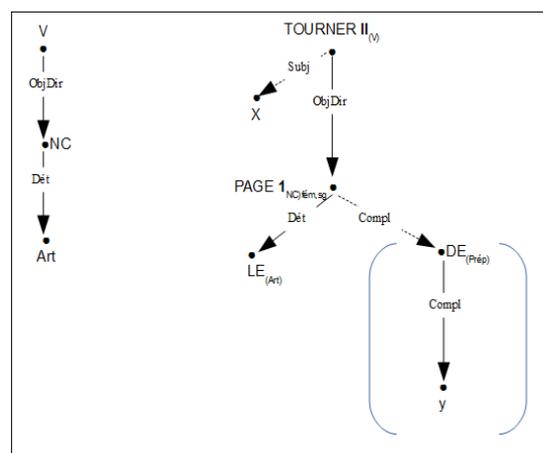


Figure 2 : Attribution d'une structure lexico-syntaxique à 'TOURNER LA PAGE'

L'identification de ces unités va permettre de récupérer toutes leurs propriétés combinatoires. Ainsi, même si la structure actancielle de base de 'TOURNER LA PAGE' est *X tourne la page*, il ne sera pas étonnant de trouver des énoncés comme (6d) dans lesquels *page* est pourvu d'un complément introduit par la préposition *de*, conformément à la structure actancielle de la lexie PAGE<sup>1</sup> : *page de X* [la page d'un livre] :

<i>tourner la page</i>	‘TOURNER LA PAGE’
tourner la page du livre	<b>tourner la page</b> <u>de la décennie de guerre</u> qui a marqué le pays

Nous verrons un peu plus loin comment il est possible de modéliser cette flexibilité des locutions.

### Identification des constituants lexicaux : entre diachronie et synchronie

Comme nous venons de le voir, identifier les constituants lexicaux des locutions participe à la modélisation de leur combinatoire, tributaire de leur statut syntagmatique. Parfois, l’identification de ces constituants peut poser question. Prenons l’exemple de la locution ‘COURIR SUR LE HARICOT’ [Jeanne commence à **courir sur le haricot** de son frère].

- (7) Finalement, ça n’est pas une vie que de bouger tout le temps. Maintenant ce sont ces déplacements incessants qui l’ennuient. Il en a pris sa dose. Ça commence à **lui courir sur le haricot**. [Frantext ; DEVILLE Patrick, *Peste & Choléra*, 2012, p. 135]

Ici nous voyons le second actant de la locution cliticisé, ce qui n’est pas sans rappeler l’emploi d’autres locutions comme CASSER LES PIEDS :

- (8) Il s’était demandé quand elle viendrait **lui casser les pieds** sur un chantier – car il lui semblait bien que casser les pieds des gens était la mission qu’elle s’était fixée dans la vie [Google Books ; Nora Roberts, *Le secret des fleurs (1) - Le dabla bleu*, Flammarion]

Il s’agit ici de l’application de la règle de cliticisation associée au lien de possession inaliénable entre un quasi-prédicat et son actant qui est, bien entendu, également compatible avec les syntagmes libres :

- (9) a. *Jeanne a marché sur les orteils* de Line.  
 b. *Jeanne lui a marché sur les orteils*.  
 c. Line **a marché sur le stylo** de Jeanne  
 d. \*Line **lui a marché sur le stylo**.

On pourra se poser, avec raison, la question de la possession inaliénable d'un haricot si l'on omet un ancien sens du vocable, justement synonyme de ORTEIL<sup>1a</sup>.

- (10) René et Marmouset avançaient péniblement dans la cohue et déjà le petit protestait :  
 – J'commence à en avoir marre de me faire trépigner sur les **haricots**.  
 [MARMOUSET, *Au lion tranquille*, 1922]

Si nous voulons activer la combinatoire de cette lexie appartenant à un état antérieur de la langue pour prédire celle de COURIR SUR LE HARICOT, nous devons l'inclure dans notre ressource. Nous lui attribuons le statut lexical de *lexie inactive*.

### Définition lexicographique

La Lexicologie Explicative et Combinatoire propose un modèle systématique de définition lexicographique (Mel'čuk et Polguère, 2018). Chaque lexie – lexème ou locution – reçoit une définition composée de :

1. un *definiendum* (défini), écrit en italique, qui décrit les positions actanciennes (X, Y, Z...) contrôlées par la lexie (autrement dit, sa structure prédicative)
2. un *definiens* (définissant), qui inclut une composante centrale (genre prochain aristotélicien) et une liste de composantes périphériques (différences spécifiques) qui jouent chacune un rôle informationnel par rapport à la composante centrale.

Pour définir 'TOURNER LA PAGE' [Gustave a tourné la page à propos de cette histoire depuis longtemps], nous procédons comme suit :

<p><i>X 'tourne la page' à propos de Y :</i></p>	<p>[[● X ayant été confronté à un événement <math>\alpha</math> non souhaitable concernant un fait ou une entité Y]]          X cause que <math>\alpha</math> de Y cesse d'avoir un effet négatif sur X</p> <ul style="list-style-type: none"> <li>• en ignorant <math>\alpha</math></li> <li>• pour se préoccuper de l'avenir de X</li> </ul>
--	--

La composante périphérique qui apparaît entre [[...]] est une composante présuppositionnelle : si elle est fausse, la situation dénotée par 'TOURNER LA PAGE' ne peut pas être vraie.

La définition lexicographique d'une lexie est la forme textuelle de l'équivalence entre les réseaux sémantiques de son *definiendum* et de son *definiens*. La figure 3 montre un extrait de cette équivalence, entre le *definiendum* (à gauche), et la composante centrale du *definiens* (à droite).

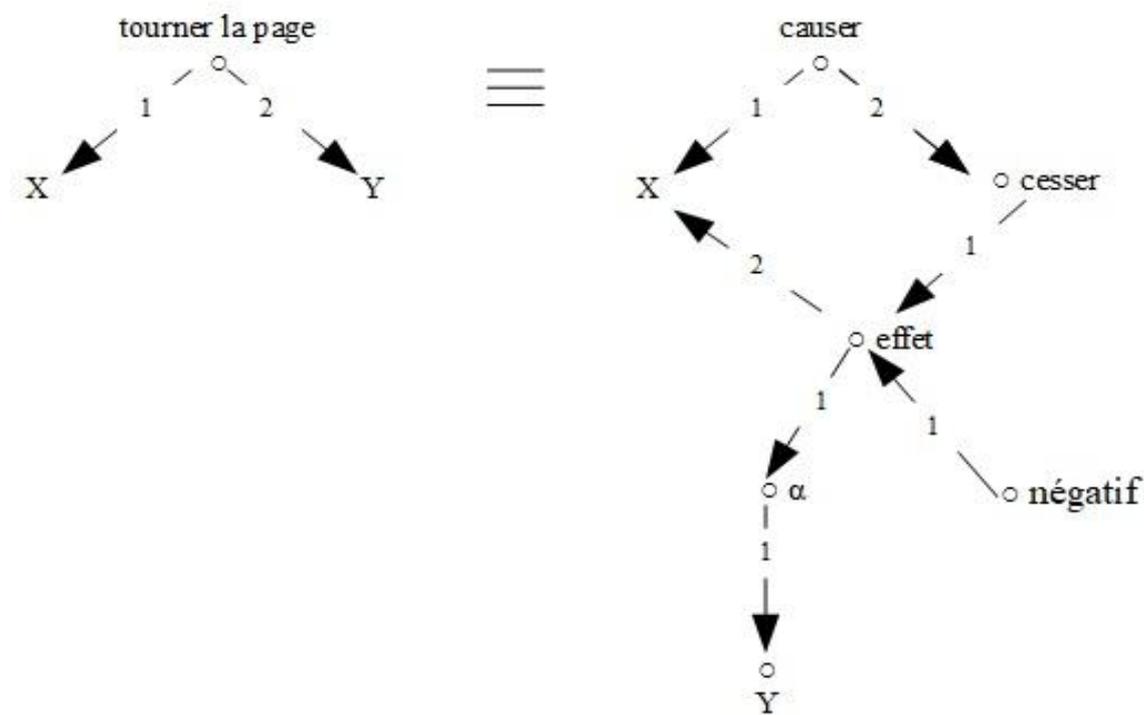


Figure 3 : Vue simplifiée des réseaux sémantiques formant la définition de tourner la page.

Ces réseaux sémantiques peuvent permettre, en lien avec les structures lexico-syntaxiques de certaines locutions, de prédire leur combinatoire.

### Intersection entre sens et structure lexico-syntaxique

Pausé (2017) a proposé la notion de *projection structurale* pour modéliser le parallèle qui existe entre le sens de certaines locutions et leur structure lexico-syntaxique. Cette modélisation fait appel à la notion de fonction lexicale présentée *supra*.

Rappelons que l'on peut modéliser le lien entre TOURNER et PAGE au moyen de la fonction lexicale **LiquFacto** (*page*) = *tourner*, qui indique que

*tourner* est le verbe utilisé pour signifier ‘rendre la page non fonctionnelle’. Il existe un lien métaphorique entre *tourner la page* et ‘TOURNER LA PAGE’, basé sur l’analogie entre le sens de la collocation et celui de la locution :

X rend α non fonctionnel = X cause que α de Y cesse d’avoir un effet négatif sur X.

Cet appariement entre une collocation et une partie ou l’intégralité de la définition d’une locution est appelé *projection structurale* :

*La projection structurale entre la SLS d’une locution et un réseau sémantique est la mise en correspondance de tout ou partie de ses constituants lexicaux avec des constituants du réseau. La projection structurale déclenche une activation de tout ou partie des propriétés combinatoires des constituants lexicaux de la locution. (Pausé 2017, 270).*

La figure 4 illustre la projection structurale entre *tourner la page* et ‘TOURNER LA PAGE’.

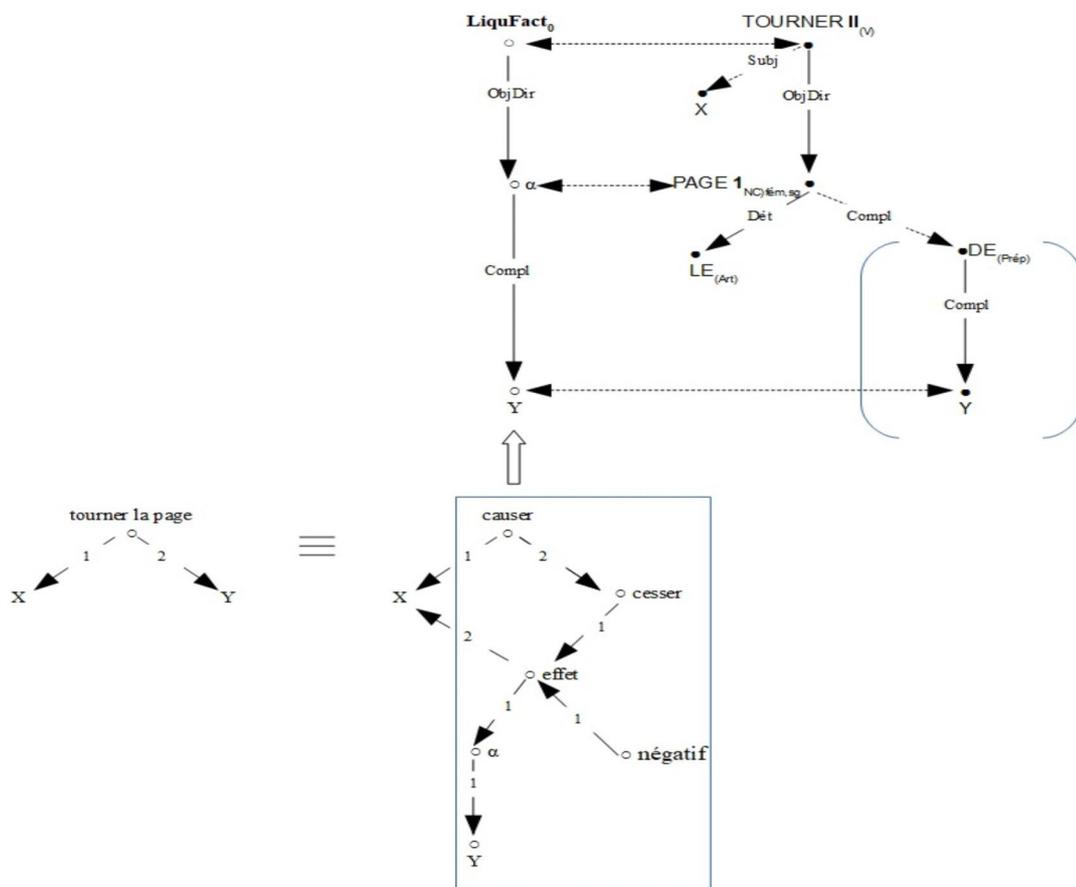


Figure 4 : Illustration de la projection structurale entre *tourner la page* et ‘TOURNER LA PAGE’.

Dans le cas de 'TOURNER LA PAGE', la projection structurale conduit à générer une lexie néologique dont le signifiant \**page*<sup>2</sup> dont le signifié est 'α'. Le syntagme locutionnel devient alors flexible en fonction de la structure communicative de l'énoncé. Ainsi, 'α' pourra être :

- le thème de l'énoncé : *La page qu'il a tournée hier [...] a eu quelque chose d'une passation des pouvoirs.* [Web, pressreader.com]
- indéfini : *Ethan espérait avoir tourné une page en quittant le service actif de la Force Mission Impossible pour un poste de formateur [...].* [FrWac]
- pourvu d'un complément qui exprime Y : *J'ai su, cette année-là, qu'un jour je tournerais la page de ma vie de prof, comme on tourne la dernière page d'un cahier de brouillon.* [FrWac]

La flexibilité formelle des locutions nous amène à nous questionner plus encore sur la frontière entre locution et collocation.

### Statut phraséologique mouvant

Certains phrasèmes peuvent changer de catégorie, oscillant entre locution et collocation. L'observation diachronique du comportement des locutions en corpus permet de constater que les constituants de certaines d'entre elles tendent à se désolidariser (Pausé et Polguère à paraître). C'est le cas de la combinaison *avoir la main verte*. Ce phrasème est souvent décrit comme une locution verbale. Néanmoins, les locuteurs emploient *main verte* avec d'autres verbes (comme *exercer*, ou *garder*), voire même sans :

- (11) a. Elle **avait** pourtant **gardé la main verte** et m'a appris qu'il faut parler aux arbres et aux plantes. [Frantext ; Maspero François, *Les abeilles et la guêpe*, 2002, 77]
- b. Venez vous creuser les méninges et **exercer votre main verte**, vous repartirez avec un panier d'idées nouvelles pour cultiver votre jardin. [FrWac]
- c. Des Brestois **à la main verte** cultivent ici fleurs ou légumes pour leurs besoins exclusifs. [FrWac]
- d. En échange d'un studio indépendant, nous cherchons une personne avec **une main verte**, pour créer et entretenir un jardin / potager. [Web, toitchezmoi.com]

On est alors amené à considérer une locution nominale associée à plusieurs collocatifs verbaux. Face à des usages de ce type, le lexicographe doit décider du statut qu'il accorde aux combinaisons phraséologiques concernées. L'appartenance à l'une ou l'autre des catégories – locution ou collocation – reflète dans un cas la volonté d'insister sur la fréquence d'apparition du phrasème dans une forme particulière, et dans l'autre la volonté de rendre compte de la combinatoire complète de l'unité.

## Conclusion

Nous espérons avoir donné au lecteur un panorama des enjeux théoriques de la description lexicographique des phrasèmes dans une ressource de type réseau lexical. La phraséologie, longtemps considérée comme un phénomène marginal, gagne aujourd'hui de plus en plus d'ampleur dans la description grammaticale des langues. Nous avons centré notre propos sur les locutions, qui manifestent des comportements qui nous amènent à questionner la frontière entre syntagmes libres et syntagmes phraséologiques. Une ressource lexicale actuelle doit absolument prendre en compte ces phénomènes et développer une description à l'interface entre sémantique et syntaxe.

Nous avons proposé la projection structurale comme moyen de modéliser la flexibilité formelle des locutions. Cette notion nous amène à questionner le modèle de définition lexicographique de la théorie Sens-Texte. Il faudra en effet l'améliorer afin qu'il puisse prendre en compte le lien entre une locution et sa structure lexico-syntaxique.

Nous avons par ailleurs peu développé la notion de phrasème non connexe, qui devra faire l'objet d'une étude à part entière.

## Références bibliographiques

- BARONI, M., BERNARDINI, S., FERRARESI, ZANCHETTA, E., The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora., *LANGUAGE RESOURCES AND EVALUATION*, 2009, **43 (3)**, 209–226.
- BLANCO, X., Inventaire lexicographique d'une sous-classe de phrasèmes délaissée : les pragmatèmes, *CAHIERS DE LEXICOLOGIE*, 2014, **104**, 133–153.
- CROFT, W., CRUSE, A., *Cognitive linguistics*, Cambridge, Cambridge University Press, 2004.
- DIDEROT, D., D'ALEMBERT, J., *Encyclopédie*, Sous la direction de André Le Breton, Laurent Durand, Antoine-Claude Briasson, et Michel-Antoine David, Paris, 1751.

- EVERT, S., *The Statistics of Word Cooccurrences. Word Pairs and Collocations*, Thèse de Doctorat, Stuttgart, Universität Stuttgart, 2005.
- FELLBAUM, C., *WordNet: An Electronic Lexical Database*, Cambridge, MA, The MIT Press, 1998.
- FELLBAUM, C., GEYKEN, A., Transforming a Corpus into a Lexical Resource: The Berlin Idiom Project, *REVUE FRANÇAISE DE LINGUISTIQUE APPLIQUÉE*, 2005, **X (2)**, 49–62.
- FILLMORE, C., Frame Semantics Meets Construction Grammar, in BERNAL, E. & DE CESARIS, J. (dir.) *Proceedings of the XIII EURALEX International Congress*, Barcelone, Institut Universitari de Lingüística Aplicada, 2008, 49-69.
- FILLMORE, C., KAY, P., O'CONNOR, M., Regularity and Idiomaticity in Grammatical Constructions: The case of « Let Alone' », *LANGUAGE*, 1988, **64**, 501–538.
- FIRTH, J.R., Modes of Meaning, in *Papers in Linguistics*, London, Oxford University Press, 1957, 190–215.
- GADER, N., LUX-POGODALLA, V., POLGUÈRE, A., Hand-Crafting a Lexical Network with a Knowledge-Based Graph Editor, in *The COLING 2012 Organizing Committee. Third Workshop on Cognitive Aspects of the Lexicon (CogAlex III)*, Mumbai, 2012, 109–125.  
<http://halshs.archives-ouvertes.fr/hal-00768798/>.
- GONZALES-REY, I., *La phraséologie du français*, Toulouse, Presses Universitaires du Midi, 2015.
- KAHANE, S., Grammaires de dépendance (Dependency grammars), *T.A.L.*, 2000, **41 (1)**.
- LANGLOTZ, A., *Idiomatic Creativity: A Cognitive-Linguistic Model of Idiom-Representation and Idiom-Variation in English*, Amsterdam/Philadelphia, John Benjamins Publishing Company, 2006.
- LUX-POGODALLA, V., POLGUÈRE, A., Construction of a French Lexical Network: Methodological Issues, in *First International Workshop on Lexical Resources, WoLeR 2011*, Ljubljana, Trojina, 2011, 55-62.
- MEL'ČUK, I., Phrasemes in Language and Phraseology in Linguistics, in EVERAERT, M., VAN DER LINDEN, E.-J., SCHENK, A., SCHREUDER, R. (dir.) *Idioms: Structural and Psychological Perspectives*, Hillsdale, Lawrence Erlbaum Associates, 1995, 167–232.
- MEL'ČUK, I., *Communicative Organization in Natural Language. The Semantic-Communicative Structure of Sentences*, Amsterdam/Philadelphia, John Benjamins, 2001.
- MEL'ČUK, I., Parties du discours et locutions, *BULLETIN DE LA SOCIÉTÉ DE LINGUISTIQUE DE PARIS*, 2006, **101 (1)**, 29–65.
- MEL'ČUK, I., Lexical Functions, in BURGER, H., DOBROVOL'SKIJ, D., KUHN, P., NORRICK, N. R. (dir.) *Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung/An international Handbook of Contemporary Research*, Sous la direction de Berlin/New-York, De Gruyter, 2007, 119–131.

- MEL'ČUK, I., Tout ce que nous voulions savoir sur les phrasèmes, mais..., *CAHIERS DE LEXICOLOGIE*, 2013, **102**, 129–149.
- 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.
- NUNBERG, G., SAG, I., WASOW, T., Idioms, *LANGUAGE*, 1994, **70 (3)**, 491–538. <https://doi.org/10.2307/416483>.
- OSHERSON, A., FELLBAUM, C., The representation of idioms in WordNet, *Proceedings Global WordNet Conference 2010, 31 January - 4 February 2010*, Mumbai, 2010.
- PAUSÉ, M.S., *Structure lexico-syntaxique des locutions du français et incidence sur leur combinatoire*, Nancy, Université de Lorraine, 2017.
- PAUSÉ, M.S., POLGUÈRE, A., Séparation phraséologique : quand les locutions s'éclatent, *CAHIERS DE LEXICOLOGIE*, à paraître, **116**, 111–49.
- PAUSÉ, M.S., SIKORA, D., Polysémie et homonymie des locutions dans un modèle lexical, in NEVEU, FR., BERGOUNIOUX, G., COTE, A.-G., FOURNIER, J.-M., HRIBA, L., PREVOST, S. (dir.) *Actes du Congrès Mondial de Linguistique Française (CMLF) 2016*, Tours, 2016.
- POLGUÈRE, A., From Writing Dictionaries to Weaving Lexical Networks, *INTERNATIONAL JOURNAL OF LEXICOGRAPHY*, 2014, **27 (4)**, 396–418. <https://doi.org/10.1093/ijl/ecu017>.
- POLGUÈRE, A., Non-compositionnalité : ce sont toujours les locutions faibles qui trinquent, *VERBUM*, 2015, **37 (2)**.
- POLGUÈRE, A., *Lexicologie et sémantique lexicale : notions fondamentales*, Troisième édition, Montréal, Les Presses de l'Université de Montréal, 2016.
- PRUVOST, J., À travers siècles, un article de dictionnaire et ses séquences figées, in ANSCOMBRE, J.-CL., MEJRI, S. (dir.) *Le figement linguistique : la parole entravée*, Paris, Honoré Champion, 2011, 341–362.
- RUPPENHOFER, J., ELLSWORTH, M., PETRUCK, M., JOHNSON, C., SCHEFFCZYK, J., *FrameNet II: Extended Theory and Practice*, Berkeley, International Computer Science Institute, 2010.
- SINCLAIR, J., *Corpus, Concordance, Collocation*, Oxford, Oxford University Press, 1981.