

Le traitement des collocations en génération de texte multilingue

Florie Lambrey François Lareau

OLST, Université de Montréal

C.P. 6128, succ. Centre-Ville

Montréal QC H3C 3J7, Canada

florie.lambrey@umontreal.ca francois.lareau@umontreal.ca

Résumé. Pour concevoir des générateurs automatiques de texte génériques qui soient facilement réutilisables d'une langue et d'une application à l'autre, il faut modéliser les principaux phénomènes linguistiques qu'on retrouve dans les langues en général. Un des phénomènes fondamentaux qui demeurent problématiques pour le TAL est celui des collocations, comme *grippe carabinée*, *peur bleue* ou *désir ardent*, où un sens (ici, l'intensité) ne s'exprime pas de la même façon selon l'unité lexicale qu'il modifie. Dans la lexicographie explicative et combinatoire, on modélise les collocations au moyen de fonctions lexicales qui correspondent à des patrons récurrents de collocations. Par exemple, les expressions mentionnées ici se décrivent au moyen de la fonction Magn : $\text{Magn}(\text{PEUR}) = \text{BLEUE}$, $\text{Magn}(\text{GRIPPE}) = \text{CARABINÉE}$, etc. Il existe des centaines de fonctions lexicales. Dans cet article, nous nous intéressons à l'implémentation d'un sous-ensemble de fonctions qui décrivent les verbes supports et certains types de modificateurs.

Abstract.

The treatment of collocations in multilingual text generation

In order to develop generic natural language generators that could be reused across languages and applications, it is necessary to model the core linguistic phenomena that one finds in language. One central phenomenon that remains problematic in NLP is collocations, such as *heavy rain*, *strong preference* or *intense flavour*, where the same idea (here, intensity), is expressed differently depending on the lexical unit it modifies. In explicative combinatory lexicography, collocations are modelled via lexical functions, which correspond to recurrent patterns of collocations. For instance, the three expressions above are all described with the function Magn : $\text{Magn}(\text{RAIN}) = \text{HEAVY}$, $\text{Magn}(\text{PREFERENCE}) = \text{STRONG}$, etc. There are hundreds of lexical functions. In this paper, we focus on the implementation of a subset of them that are used to model support verbs and some modifiers.

Mots-clés : génération automatique de texte multilingue ; collocations ; fonctions lexicales ; théorie sens-texte.

Keywords: multilingual natural language generation ; collocations ; lexical functions ; meaning-text theory.

1 Introduction

La génération automatique de texte (GAT) vise à présenter de façon automatique de l'information que l'on veut communiquer en langue naturelle. À ces fins, l'approche symbolique encode les connaissances linguistiques pertinentes dans des dictionnaires et grammaires, des ressources coûteuses à développer. La plupart des générateurs de texte sont monolingues. Si on veut pouvoir générer des textes dans plusieurs langues en même temps, il faut adapter les modèles linguistiques pour chaque nouvelle langue. Pour créer de tels systèmes de génération automatique de texte multilingue (GATM), il existe deux stratégies. La première consiste à transposer les ressources linguistiques vers la ou les langue(s) visée(s) (Alshawi & Pulman, 1992; Kim *et al.*, 2003). La seconde consiste à factoriser ce qui peut l'être afin de partager certains éléments de description entre les langues (Avgustinova & Uszkoreit, 2000; Bateman *et al.*, 2005; Lareau & Wanner, 2007; Santaholma, 2008). Notre travail se situe dans la mouvance de cette seconde approche.

Les grammaires et les dictionnaires modélisent divers phénomènes, dont la lexicalisation (Wanner, 1996a; Polguère, 1997, 2000; Reiter & Dale, 2000). Dans le cadre de la GATM, le modèle de lexicalisation doit être conçu de manière à être aussi générique que possible afin de maximiser la quantité de règles communes aux langues à traiter. Pour ce faire, il faut chercher à modéliser les phénomènes linguistiques centraux partagés par plusieurs langues et s'en servir comme base sur

laquelle s’ajoutent les modules spécifiques à chaque langue. Alors que beaucoup de sens se lexicalisent toujours à l’aide d’un même mot dans une langue donnée, d’autres vont se réaliser différemment en fonction du contexte linguistique comme le sens de causation dans *donner le rhume*, *mettre dans le pétrin* ou *jeter un sort*. Ces dépendances lexicales se traduisent par des contraintes combinatoires dans ce qu’on appelle des collocations. Les collocations sont omniprésentes dans les textes naturels mais restent problématiques en traitement automatique des langues (TAL). Malgré leur apparente diversité, les collocations peuvent se généraliser et former des patrons communs à plusieurs langues. L’objectif de cet article est de décrire la méthodologie qui a permis de modéliser certains de ces patrons et de les intégrer dans un système générique de GATM. Ce travail, encore dans sa phase initial se situe dans le cadre du projet GÉCO (GÉNÉration de COLlocations).

2 Les collocations et les fonctions lexicales en GATM

Plusieurs projets ont porté sur la création et le partage de ressources linguistiques appliquées à la génération de texte multilingue ou à la traduction automatique, notamment Boguslavsky *et al.* (2004), Bateman *et al.* (2005) et Lareau & Wanner (2007). Le processus de lexicalisation est complexe en raison de l’interdépendance entre les choix lexicaux et syntaxiques (Steinlin, 2003). Cela se traduit par des contraintes combinatoires devant être décrites dans les ressources linguistiques. Les collocations sont une parfaite illustration de ces contraintes. Par exemple, *faire un pas*, *dar un paso* (litt. ‘donner un pas’), et *take a step* (litt. ‘prendre un pas’) instancient le même phénomène linguistique (l’emploi d’un verbe support) mais se réalisent différemment d’une langue à l’autre. Ce type d’information est crucial en GATM pour simuler au maximum le langage naturel.

Les collocations, comme *peur bleue*, *brouillard dense* et *remercier chaleureusement*, sont toujours composées d’une base et d’un collocatif, et le choix du collocatif est déterminé par celui de la base. Même si cette préférence lexicale est arbitraire, il est possible d’en tirer une généralisation sous forme de patron. En l’occurrence, dans les exemples donnés ici, on retrouve un même patron d’intensification, que l’on peut représenter sous forme de ratio : $\frac{\text{bleue}}{\text{peur}} = \frac{\text{dense}}{\text{brouillard}} = \frac{\text{chaleureusement}}{\text{remercier}}$. Ce genre de patron se retrouve à travers les langues et correspond à ce qu’on appelle des fonctions lexicales (FL) (Mel’čuk, 1995; Wanner, 1996b; Kahane & Polguère, 2001; Apresjan *et al.*, 2002); dans ce cas-ci, il s’agit de la fonction Magn . Dans notre système, les collocations sont décrites à l’aide de telles fonctions.

Les FL ont déjà été utilisées en GAT (monolingue et multilingue). Heid & Raab (1989), Apresjan *et al.* (2002) et Iordanskaja *et al.* (1996), entre autres, les utilisaient dans le cadre de modèles basés sur la théorie Sens-Texte (TST), alors que Steinlin (2003) les employait dans le cadre de la grammaire d’arbres adjoints (TAG), et Lareau *et al.* (2011, 2012) les implémentaient en grammaire lexicale fonctionnelle (LFG). Enfin, (Fontenelle, 1997) employait les FL en vue de la création d’une base de donnée sémantique-lexicale et faisait référence à d’autres travaux portant sur l’utilisation des FL pour la GAT et en TAL de manière générale (Van der Wouden, 1992; Iordanskaja *et al.*, 1996; Lee & Evens, 1996). Ces implémentations, toutefois, étaient partielles : elles traitaient un sous-ensemble de FL dans un domaine particulier. Un des objectifs du projet GÉCO est d’implémenter de façon exhaustive les FL afin de décrire les patrons de collocations récurrents en langue générale. Dès lors, l’encodage des FL s’oriente autour d’une structuration modulaire de notre système, nous amenant à opérer des regroupements de FL en patrons plus généraux. Le résultat de cette implémentation sera donc un outil répertoriant toutes ces FL et libre d’accès. Le choix d’utiliser les FL pour la GATM s’est imposé car, comme l’expliquent Lareau & Wanner (2007), elles permettent de réduire le nombre de règles spécifiques aux langues. Comme l’a montré Fontenelle (1997), les FL standards permettent d’encoder des distinctions sémantiques fines, comme celle entre les verbes anglais $\text{RAISE} (= \text{CausPredPlus}(\text{PRICE}))$ et $\text{RISE} (= \text{IncepPredPlus}(\text{PRICE}))$. Il arrive toutefois que les FL manquent de granularité ; par exemple, *recommander fortement* et *recommander chaleureusement* sont deux valeurs possibles de la même FL $\text{Magn}(\text{RECOMMENDER})$, même si leur sens n’est pas tout à fait identique. Par ailleurs, il n’existe pas de consensus solide sur l’inventaire des FL standards (Fontenelle, 1997; Jousse, 2003, 2010).

3 Le système GÉCO

Le système de GATM proposé ici se base sur la grammaire MARQUIS (Lareau & Wanner, 2007; Wanner *et al.*, 2010), implémentée sur la plateforme MATE. MATE est un transducteur de graphes libre conçu par Bohnet & Wanner (2010) et propose un environnement pour le développement de grammaires et dictionnaires. Conçu à la base pour modéliser l’approche multistratale de la TST, il peut manipuler divers types de graphes et construit, à partir d’un graphe source, un graphe correspondant d’un niveau supérieur, sans modifier la structure de départ. Le processus de génération s’articule autour de plusieurs niveaux : conceptuel, sémantique, syntaxique profond et de surface, morphologique profond et de

surface, et phonologique. Des règles de correspondance permettent de passer d'un niveau à l'autre. Étant donné que le travail présent se focalise sur l'étape de lexicalisation en vue de décrire les patrons de collocations récurrents, nous nous sommes focalisés sur l'interface entre la structure sémantique, un graphe orienté acyclique (GOA), et syntaxiques profonde et de surface, des arbres de dépendance non linéarisés. Le niveau sémantique représente le message que l'on souhaite communiquer alors que le niveau syntaxique prend en charge la structure de la phrase. Les règles de correspondance entre niveaux sont développées de manière tripartite incluant une structure source, une structure cible et un ensemble de conditions d'application de la règle. Le modèle est transductif et non transformationnel, c'est-à-dire qu'on ne modifie jamais la structure donnée en entrée, mais on construit une ou plusieurs nouvelle(s) structure(s) d'un niveau de représentation adjacent. La mise en correspondance entre les structures n'est pas déterministe : à une structure donnée en entrée correspondent en général plusieurs structures. En d'autres termes, le paraphrasage fait partie intégrante du système.

4 Implémentation des patrons de collocations

4.1 Sélection de patrons pertinents

La grammaire de MARQUIS (Lareau & Wanner, 2007) établit des règles de l'interface sémantique-syntaxe très générales et capables de traiter six langues européennes avec un faible nombre de règles spécifiques à une langue donnée. Ils ont montré que cette architecture basée sur des dictionnaires riches et des règles aussi génériques que possible accélérerait grandement leur travail. C'est dans cette optique d'optimisation que nous avons généralisé encore plus ces règles dans notre système de GATM à l'aide de patrons de collocations. La création de patrons généraux, des classes de FL partageant des propriétés similaires, s'insère dans une perspective de développement continu de notre plateforme et vise à faciliter l'implémentation de nouvelles FL en plus du maintien des structures existantes.

Les premiers patrons implémentés modélisent la lexicalisation des verbes supports et de certains modificateurs. Ces phénomènes sont très fréquents et ont un impact important sur la structure syntaxique de la phrase. Ils sont décrits au moyen de FL standard simples. Ce sont donc ces FL que nous avons modélisées pour le moment en accord avec les descriptions communément acceptées de ces FL (Mel'čuk *et al.*, 1995; Fontenelle, 1997; Mel'čuk, 2004). C'est pour cela que les regroupements établis jusqu'ici se basent essentiellement sur les propriétés syntaxiques des FL traitées. Par exemple, le patron des verbes supports correspond toujours à un collocatif verbal vide de sens et sa valence. D'autres patrons sont définis en fonction de la relation syntaxique entretenue entre la base et le collocatif ; c'est le cas du patron *modificateur*. Par ailleurs, nos patrons modélisent les informations d'ordre sémantique contenues dans les FL, ce qui se reflète à l'aide d'indices faisant référence aux actants de la base. Par exemple, A_2 retourne une expression adjectivale qui contient la base et qui modifie le deuxième actant de la base. En résumé, l'élaboration des patrons de collocations s'est faite dans l'optique de pouvoir faire des regroupements de comportements syntaxiques similaires. Chaque patron est incarné par une règle de lexicalisation dans MATE. Le projet GÉCO prévoit d'implémenter au fur et à mesure tous les types de relations décrits par les FL standard simples, complexes et non-standard. Les patrons implémentés sont présentés dans la table 1.

Le principal risque de cette approche est d'opérer des regroupements inappropriés de FL menant à un manque de granularité général du système. Par ailleurs, la formulation des règles ainsi que la structuration des dictionnaires auxquels elles font référence est contrainte par l'environnement MATE. L'encodage des FL est également dépendant de cette plateforme.

4.2 Implémentation des verbes supports

Les verbes supports ont été bien décrits dans la TST, notamment par Wanner (1996b); Fontenelle (1997); Apresjan *et al.* (2002); Mel'čuk (2004) entre autres. Il existe plusieurs caractéristiques partagées par ce type de verbes, en particulier le fait qu'ils sont sémantiquement vides (dans leur usage en tant que collocatif). Un verbe support n'est donc pas présent dans la structure sémantique de l'énoncé, mais est construit par des règles de la grammaire afin de satisfaire les contraintes syntaxiques et lexicales de la langue. Il en existe trois types majeurs définis en fonction du rôle syntaxique de la base vis-à-vis du verbe support : *Oper* (la base est l'objet direct du collocatif), *Func* (la base est le sujet syntaxique) et *Labor* (la base est l'objet indirect). Ces FL peuvent posséder un indice qui donne la position syntaxique des actants de la base par rapport au verbe support¹. Ainsi, DONNER est l' $Oper_1$ de GIFLE car, d'une part, le sujet syntaxique du verbe support correspond au premier actant de la base et, d'autre part, la base est l'objet direct du verbe support (cf. Mel'čuk 2004).

1. Comme l'indique un relecteur, selon la définition standard des FL les indices font référence aux actants syntaxiques profonds du prédicat. Cependant, baser nos règles sur les actants sémantiques facilite le processus de création de règles.

Patron	Fonction lexicale	Exemple
Modificateur	Magn	<i>peur bleue</i>
	Ver	<i>argument convaincant</i>
Verbe support	Func ₀	<i>la neige tombe</i>
	Func ₁	<i>la faute incombe à X</i>
	Func ₃	<i>la discussion porte sur Z</i>
	Oper ₁	<i>X essuie un échec</i>
	Oper ₂	<i>Y reçoit un coup</i>
	Oper ₁₂	<i>X fait un compliment à Y</i>
	Labor ₁₂	<i>X soumet Y à l'analyse</i>
Nom pléonastique	Figur	<i>rideau de fumée</i>
Préposition	Loc _{ad}	<i>au front</i>
	Loc _{in}	<i>au sein du personnel</i>
Adjectivisation/Adverbialisation	A ₂	<i>(Y est) couvert de mépris</i>
	Adv ₁	<i>(X fait quelque chose) avec joie</i>

TABLE 1 – Patrons de fonctions lexicales implémentés

Dans la grammaire MARQUIS, à chaque FL décrivant un verbe support correspond une règle de lexicalisation prenant en compte le type de verbe support ainsi que le nombre et le type d’actants réalisés. Il y a en tout six règles pour les verbes supports dans MARQUIS : Func₀, Func_i, Oper₀, Oper_i, Oper_{ij} et Labor_{ij} (chacune de ces règles possède une illustration linguistique dans la table 1). La figure 1 illustre schématiquement la règle Oper_i qui permet de générer des énoncés comme *essuyer un échec*, *donner une baffe* ou *faire un pas*. La partie gauche contient deux nœuds, X et A ; X est un sens prédicatif et A est son i-ème argument, ce qui est représenté par la flèche partant de X vers A, symbolisant le fait que X domine sémantiquement A. Par exemple, pour *Paul essuie un échec*, le sens est ‘échouer(Paul)’, où X=‘échouer’, A=‘Paul’ et i=1 (puisque ‘Paul’ est le premier argument de ‘échouer’). La partie droite de Oper_i donne la structure syntaxique profonde correspondante sous forme d’arbre de dépendance. Elle génère trois nœuds dans la structure syntaxique : Oper_i, C et “lex(X)”. Le nœud C est la réalisation de A (ce que les pointillés indiquent). Cette règle se contente de créer dans la structure syntaxique le nœud correspondant à A dans la structure sémantique mais laisse le soin à une autre règle d’opérer sa lexicalisation. Le sens prédicatif (dans notre exemple, ‘échouer’) peut s’exprimer par le verbe ESSUYER accompagné du nom ÉCHEC ; il a donc deux nœuds correspondants en syntaxe : le verbe Oper_i et le nom qui va lexicaliser X (“lex(X)”). La lexicalisation de X, représentée ici par “lex(X)”, est récupérée dans un dictionnaire sémantique contenant l’instruction “lex(échouer) = échec” (et potentiellement d’autres lexicalisations pour ce sens, comme ÉCHOUER). La valeur du collocatif est ensuite calculée lors d’une deuxième étape de lexicalisation, alors qu’elle sera récupérée dans un dictionnaire où on trouve l’information “Oper₁(échec) = essuyer” (et potentiellement d’autres collocatifs, comme SUBIR). La structure syntaxique est figée : la base est forcément l’actant syntaxique II (objet direct) et l’argument sémantique de la base est toujours l’actant I (sujet).

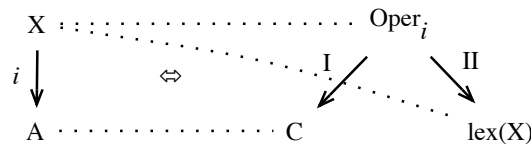
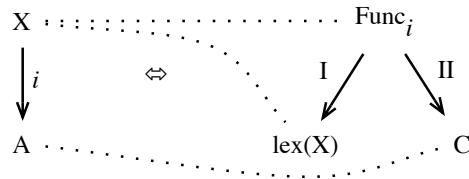


FIGURE 1 – Oper_i dans MARQUIS

Dans notre système, nous avons réduit le nombre de règles pour les verbes supports à trois. Nos règles sont donc encore plus générales que celles de Lareau & Wanner (2007). La figure 4 montre la règle qui traite à la fois Func₁, Func₂, Func₃, etc., et Oper₁, Oper₂, Oper₃, etc. La différence majeure porte sur la description des liens syntaxiques entre le nœud f et ses actants syntaxiques profonds. Nos règles exploitent au maximum la structuration des différents dictionnaires. Nous utilisons un dictionnaire LF où nous stockons des entrées comme celle de Oper₁ dans la figure 3. Cette entrée encode toutes les informations requises pour créer une structure syntaxique profonde correcte à l’aide de Oper₁ comme sa partie du discours et son régime (“gp”, pour *government pattern*). On apprend donc que la base, X dans notre exemple,

FIGURE 2 – $Func_i$ dans MARQUIS

```

Oper1 {
  dpos = V
  gp = { base = II
        1 = I
        I = { dpos = N
              rel = subj } } }

```

FIGURE 3 – $Oper_1$ dans le dictionnaire *LF*

se réalisera nécessairement comme l'actant syntaxique **II** pour chaque $Oper_1$. La relation sémantique **1** correspond à la relation syntaxique **I**. Par ailleurs, cette entrée présente les contraintes syntaxiques imposées sur l'actant **I** de $Oper_1$ (ce doit être un nom introduit par la relation syntaxique de surface **subj**). Dans la figure 4, les instructions “*lf.f.i*” et “*lf.f.base*” sont des chemins permettant de récupérer dans le dictionnaire *LF* les relations syntaxiques relatives (**I** ou **II**) de la base et du premier actant par rapport au verbe support. La FL représentant le verbe support est symbolisée par la variable *f*. L'intérêt de cette règle et qu'elle permet de faire abstraction de la nature exacte de la FL et fonctionne comme un modèle de lexicalisation plus général encore. Cette règle commence par chercher dans le dictionnaire *lexicon*, de façon non-déterministe, l'entrée correspondant à “*lex(X)*” et récupère une FL qui respecte les contraintes de la règle en figure 4. Dans un second temps elle enregistre le nom de la FL dans la variable *f*, à laquelle on peut ensuite faire appel dans la règle. Grâce au dictionnaire *LF*, on peut récupérer la structure syntaxique associée à la FL en question. Cette même règle décrit ainsi les FL $Oper_1, Oper_2, Oper_3, etc.$, et $Func_1, Func_2, Func_3, etc.$

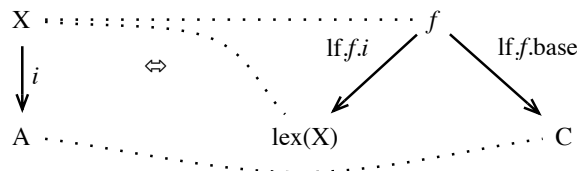


FIGURE 4 – Verbes supports à deux actants syntaxiques dans GÉCO

Nous avons créé des règles comparables pour les patrons illustrés dans la table 1. En plus de ces règles, des ressources lexicographiques sont en cours d'élaboration, notamment un dictionnaire des fonctions lexicales (*LF*), un dictionnaire d'unités sémantiques (*semanticon*) et un dictionnaire contenant les informations relatives à la combinatoire des unités lexicales (*lexicon*) dont la figure 5 illustre une entrée. Lors du processus de lexicalisation, les règles activent l'information

```

échec {
  dpos = N
  gp = { 1 = I
        I = { dpos = N } }
  lf = { name = Oper1
        value = subir }
  lf = { name = Oper1
        value = essayer } }

```

FIGURE 5 – Entrée pour ÉCHEC dans le dictionnaire *lexicon*

contenue dans chacun de ces dictionnaires. Il est donc nécessaire de développer des dictionnaires détaillés fournissant une

analyse fine des unités lexicales et d'optimiser leur structuration pour accéder facilement à ces données. La spécification exhaustive de la combinatoire des unités lexicales offre également l'avantage d'encoder des relations d'équivalence syntaxique. La figure 5 montre que le lexème ÉCHEC possède deux FL dénotant des verbes supports quasiment équivalents. Les expressions *subir un échec* et *essayer un échec* peuvent toutes deux être générées à partir d'une structure sémantique unique et ainsi rendre les règles non déterministes. Ainsi, ce système est capable de générer des paraphrases au sein d'une même langue. C'est également ce qui rend notre système multilingue. L'élaboration de tels dictionnaires est évidemment coûteuse en temps et requiert un haut niveau d'expertise en lexicographie. Nous nous servons de ressources existantes, notamment le *Réseau Lexical du Français* (Polguère, 2014), le *DicoEnviro* (L'Homme & Laneville, 2009), le *DicoInfo* (L'Homme, 2009), ainsi que le *DiCE* (Alonso Ramos, 2003).

5 Conclusion

Les fonctions lexicales sont intéressantes pour la GATM car elles permettent de modéliser des phénomènes linguistiques précis et récurrents d'une langue à l'autre, comme la gestion de paraphrases et la lexicalisation complexe. Les patrons modélisés jusqu'ici correspondent à des FL standard simples de base. Plusieurs autres types de FL viendront s'ajouter à l'avenir, notamment avec la modélisation des verbes de réalisation ($Real_i$, $Fact_i$, ...), les FL complexes, et éventuellement les FL non standard. La version finale de notre outil couvrira donc un ensemble conséquent de FL et sera disponible pour toute personne cherchant à travailler avec les FL.

Un des objectifs de notre système est de rendre utilisables pour la GAT des ressources lexicales électroniques existantes qui se servent des FL, comme celles mentionnées ci-dessus. Lors de l'implémentation des dictionnaires à partir de ces ressources, 5% des entrées lexicales existantes seront mises de côté en vue d'effectuer une évaluation de GÉCO. Il sera donc possible de mesurer la qualité de nos patrons à l'aide des mesures de précision et de rappel.

Références

- ALONSO RAMOS M. (2003). Hacia un diccionario de colocaciones del español y su codificación. *Lexicografía computacional y semántica*, **64**, 11–34.
- ALSHAWI H. & PULMAN S. G. (1992). Ellipsis, Comparatives, and Generation. In H. ALSHAWI, Ed., *The Core language engine*, chapter 13, p. 251–275. Cambridge, Massachusetts and London, England : The MIT Press.
- APRESJAN J. D., BOGUSLAVSKY I. M., IOMDIN L. L. & TSINMAN L. L. (2002). Lexical functions in actual NLP applications. In *Computational Linguistics for the New Millennium : Divergence or Synergy ? Festschrift in Honour of Peter Hellwig on the occasion of his 60th Birthday*, p. 55–72. Frankfurt : Peter Lang.
- AVGUSTINOVA T. & USZKOREIT H. (2000). An ontology of systemic relations for a shared grammar of slavic. In *Proceedings of Coling 2000*, Saarbrücken.
- BATEMAN J. A., KRUIJFF-KORBAYOVÁ I. & KRUIJFF G.-J. (2005). Multilingual resource sharing across both related and unrelated languages : An implemented, open-source framework for practical natural language generation. *Research on Language and Computation*, **15**, 1–29.
- BOGUSLAVSKY I., IOMDIN L. & SIZOV V. (2004). Multilinguality in ETAP-3 : Reuse of lexical resources. In *Proceedings of the Workshop on Multilingual Linguistic Resources*.
- BOHNET B. & WANNER L. (2010). Open source graph transducer interpreter and grammar development environment. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10)*, p. 211–218, Malta.
- FONTENELLE T. (1997). *Turning a bilingual dictionary into a lexical-semantic database*. Tübingen : Max Niemeyer Verlag.
- HEID U. & RAAB S. (1989). Collocations in multilingual generation. In *Proceedings of the fourth conference on European chapter of the Association for Computational Linguistics (EACL'89)*, p. 130–136, Manchester.
- IORDANSKAJA L., KIM M. & POLGUÈRE A. (1996). Some Procedural Problems in the Implementation of Lexical Functions for Text Generation. In (Wanner, 1996b), p. 279–297.
- JOUSSE A.-L. (2003). Normalisation des fonctions lexicales. Mémoire de DEA, Université Paris 7.
- JOUSSE A.-L. (2010). *Modèle de structuration des relations lexicales basé sur le formalisme des fonctions lexicales*. Thèse de doctorat, Université de Montréal / Université Paris 7.

- KAHANE S. & POLGUÈRE A. (2001). Formal foundation of lexical functions. In *Proceedings of the Workshop on Collocations at ACL 2001*, Toulouse.
- KIM R., DALRYMPLE M., KAPLAN R. M., KING T. H., MASUICHI H. & OHKUMA T. (2003). Multilingual grammar development via grammar porting. In *ESSLLI 2003 Workshop on Ideas and Strategies for Multilingual Grammar Development*.
- LAREAU F., DRAS M., BÖRSCHINGER B. & DALE R. (2011). Collocations in multilingual natural language generation : Lexical functions meet lexical functional grammar. In D. MOLLÁ & D. MARTINEZ, Eds., *Proceedings of the Australasian Language Technology Association Workshop*, p. 95–104, Canberra.
- LAREAU F., DRAS M., BÖRSCHINGER B. & TURPIN M. (2012). Implementing lexical functions in XLE. In M. BUTT & T. H. KING, Eds., *Proceedings of LFG'12*, p. 362–382, Denpasar, Indonesia : CSLI Publications.
- LAREAU F. & WANNER L. (2007). Towards a generic multilingual dependency grammar for text generation. In T. H. KING & E. M. BENDER, Eds., *Proceedings of the GEAF07 Workshop*, p. 203–223, Stanford : CSLI.
- LEE W. & EVENS M. (1996). Generating cohesive text using lexical functions. In (Wanner, 1996b), p. 299–306.
- L'HOMME M.-C. (2009). *DiCoInfo. Le dictionnaire fondamental de l'informatique et de l'Internet*. OLST, Université de Montréal.
- L'HOMME M.-C. & LANEVILLE M.-E. (2009). *DiCoEnviro. Le dictionnaire fondamental de l'environnement*. OLST, Université de Montréal.
- MEL'ČUK I. A. (1995). The future of the lexicon in linguistic description and the explanatory combinatorial dictionary. In I.-H. LEE, Ed., *Linguistics in the morning calm*, volume 3. Seoul : Hanshin.
- MEL'ČUK I. A. (2004). Verbes supports sans peine. *Linguisticae Investigationes*, **27**(2), 203–217.
- MEL'ČUK I. A., CLAS A. & POLGUÈRE A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Universités francophones. Louvain-la-Neuve : Duculot.
- POLGUÈRE A. (1997). Engineering text generation. *La Tribune des industries de la langue et de l'information électronique*, **23-24**, 21–39.
- POLGUÈRE A. (2000). A “natural” lexicalization model for language generation. In *Proceedings of SNLP 2000*, p. 37–50, Chiangmai.
- POLGUÈRE A. (2014). From writing dictionaries to weaving lexical networks. *International Journal of Lexicography*, **27**(4), 396–418.
- REITER E. & DALE R. (2000). *Building Natural Language Generation Systems*. Cambridge : Cambridge University Press.
- SANTAHOLMA M. (2008). Multilingual grammar resources in multilingual application development. In *Coling 2008 : Proceedings of the workshop on Grammar Engineering Across Frameworks*, p. 25–32, Manchester.
- STEINLIN J. (2003). Générer des collocations. Mémoire de DEA, Université Paris 7.
- VAN DER WOUDE T. (1992). Prolegomena to a multilingual description of collocations. In H. TOMMOLA & K. VARANTOLA, Eds., *Proceedings of EURALEX 1992*, p. 449–456, Tampere.
- WANNER L. (1996a). Lexical Choice in Text Generation and Machine Translation. *Machine Translation*, **11**(1–3).
- L. WANNER, Ed. (1996b). *Lexical functions in lexicography and natural language processing*, volume 31 of *Studies in language companion series*. Amsterdam/Philadelphia : John Benjamins.
- WANNER L., BOHNET B., BOUAYAD-AGHA N., LAREAU F. & NICKLASS D. (2010). MARQUIS : Generation of user-tailored multilingual air quality bulletins. *Applied Artificial Intelligence*, **24**(10), 914–952.